

**Proceedings of the  
Tenth IEEE Workshop on  
Statistical Signal and Array Processing**



---

*Sponsored by*  
**The IEEE Signal Processing Society**

---

**August 14-16, 2000**  
**Pocono Manor Inn, Pocono Manor, Pennsylvania, USA**



**20001024 007**

---

*Supported by*  
**Office of Naval Research      Air Force Research Laboratory      Villanova University**  
**USA                                      USA                                      USA**

**DTIC QUALITY INSPECTED 4**

# REPORT DOCUMENTATION PAGE

Form Approved  
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Washington Headquarters Service, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington, DC 20503.

PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.

1. REPORT DATE (DD-MM-YYYY) 15-10-2000		2. REPORT DATE Final		3. DATES COVERED (From - To) January 2000 – September 2000	
4. TITLE AND SUBTITLE  THE 10TH IEEE SIGNAL PROCESSING WORKSHOP ON STATISTICAL SIGNAL AND ARRAY PROCESSING				5a. CONTRACT NUMBER	
				5b. GRANT NUMBER N00014-00-1-0014	
				5c. PROGRAM ELEMENT NUMBER	
6. AUTHOR(S)  Moeness G. Amin				5d. PROJECT NUMBER	
				5e. TASK NUMBER	
				5f. WORK UNIT NUMBER	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Villanova University 800 Lancaster Ave Villanova, PA 19085				8. PERFORMING ORGANIZATION REPORT NUMBER  Acc: 527639	
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Office of Naval Research, Program Officer: W. Miceli Ballston Center Tower One 800 North Quincy Street Arlington, VA 22217-5660				10. SPONSOR/MONITOR'S ACRONYM(S)	
				11. SPONSORING/MONITORING AGENCY REPORT NUMBER	
12. DISTRIBUTION AVAILABILITY STATEMENT  Approved for Public Release; Distribution is Unlimited					
13. SUPPLEMENTARY NOTES					
14. ABSTRACT  This is the Proceedings of the 10th IEEE Workshop on Statistical Signal and Array Processing (SSAP), which was held at the Pocono Manor Inn, Pocono Manor, Pa during the period of August 14th-16th, 2000. The Workshop featured four keynote speakers whose talks covered the areas of Radar and Sonar Signal Processing; Time-Delay Estimation; Space-Time Codes; and Multi-carrier CDMA. The Workshop offered traditional and new research topics. It included one session on Radar Signal Processing, one session on Signal Processing for GPS, one session on Network Traffic Modeling, one session on Statistical Signal Processing, one session on Acoustical Signal Processing, two sessions on Time-Frequency Analysis, two sessions on Array Processing, three sessions on Second and Higher Order Statistics, and four sessions on Signal Processing for Communications. The workshop received the highest number of paper submissions compared to previous workshops in the same area, and the technical committee carefully selected high quality papers for presentations. The 2000 IEEE-SSAP Workshop was a tremendous success in all aspects.					
15. SUBJECT TERMS Radar Signal Processing, Statistical Signal Processing, Signal Processing for Communications, Time-Frequency Analysis, Array Processing					
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			19b. TELEPHONE NUMBER (Include area code)
U	U	U	UU	736	



# **Proceedings of the Tenth IEEE Workshop on Statistical Signal and Array Processing**

---

*Sponsored by*  
**The IEEE Signal Processing Society**

---

**August 14-16, 2000**  
**Pocono Manor Inn, Pocono Manor, Pennsylvania, USA**

---

*Supported by*

<b>Office of Naval Research</b>	<b>Air Force Research Laboratory</b>	<b>Villanova University</b>
<b>USA</b>	<b>USA</b>	<b>USA</b>

## **Proceedings of the Tenth IEEE Workshop on Statistical Signal and Array Processing**

Copyright and Reprint Permission: Abstracting is permitted with credit to the source. Libraries are permitted to photocopy beyond the limit of U.S. copyright law for private use of patrons those articles in this volume that carry a code at the bottom of the first page, provided the per-copy fee indicated in the code is paid through Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923. For other copying, reprint or republication permission, write to IEEE Copyrights Manager, IEEE Operations Center, 445 Hoes Lane, P.O. Box 1331, Piscataway, NJ 08855-1331. All rights reserved. Copyright © 2000 by The Institute of Electrical and Electronics Engineers, Inc.

IEEE Catalog Number: 00TH8496

ISBN: 0-7803-5988-7 (hardbound)

Library of Congress Number: 99-69422

---

## IEEE SSAP-2000 Workshop Committee

---

### *General and Organizational Chair*

Moeness Amin  
Villanova University, USA.  
e-mail:moeness@ece.vill.edu

### *Technical Chair*

Mike Zoltowski  
Purdue University, USA  
e-mail:mikedz@ecn.purdue.edu

### *Finance*

Kevin Buckley  
Villanova University, USA  
e-mail:buckley@ece.vill.edu

### *Publicity*

Rick Blum  
Lehigh University, USA  
e-mail:rblum@EECS.Lehigh.EDU

### *Proceedings*

Bill Jemison  
Lafayette College, USA  
e-mail: w.d.jemison@ieee.org

### *Local Arrangement*

Wojtek Berger  
University of Scranton, USA  
e-mail:Berger@Scranton.edu

### *Asian Liaison*

Rahim Leyman  
e-mail:EARLEYMAN@ntu.edu.sg

### *Australian Liaison*

Abdelhak Zoubir  
e-mail:zoubir@mail.atr.curtin.edu.au

### *European Liaison*

Pierre Comon  
e-mail:Pierre.Comon@i3s.unice.fr

# Table of Contents

## Session MA-1. SIGNAL PROCESSING FOR COMMUNICATIONS I

<b>Multistage Multiuser Detection for CDMA with Space-Time Coding</b> <i>Y. Zhang and R. S. Blum — Lehigh University</i>	1
<b>Adaptive MAP Multi-User Detection for Fading CDMA Channels</b> <i>C. Andrieu and A. Doucet — Cambridge University, UK</i> <i>A. Touzni — NxtWave Communications</i>	6
<b>Analysis of a Subspace Channel Estimation Technique for Multicarrier CDMA Systems</b> <i>C. J. Escudero, D. I. Iglesia, M. F. Bugallo, and L. Castedo — Universidad de La Coruña, Spain</i>	10
<b>Blind Adaptive Asynchronous CDMA Multiuser Detector Using Prediction Least Mean Kurtosis Algorithm</b> <i>K. Wang and Y. Bar-Ness — New Jersey Institute of Technology</i>	15
<b>MMSE Equalization for Forward Link in 3G CDMA: Symbol-Level Versus Chip-Level</b> <i>T. P. Krauss, W. J. Hillery, and M. D. Zoltowski — Purdue University</i>	18
<b>Transform Domain Array Processing for CDMA Systems</b> <i>Y. Zhang and M. G. Amin — Villanova University</i> <i>K. Yang — ATR Adaptive Communications Research Laboratories, Japan</i>	23
<b>Sectorized Space-Time Adaptive Processing for CDMA Systems</b> <i>K. Yang, Y. Mizuguchi — ATR Adaptive Communications Research Laboratories, Japan</i> <i>Y. Zhang — Villanova University</i>	28
<b>Demodulation of Amplitude Modulated Signals in the Presence of Multipath</b> <i>Z. Xu and P. Liu — University of California</i>	33
<b>Multichannel and Block Based Precoding Methods for Fixed Point Equalization of Nonlinear Communication Channels</b> <i>A. J. Redfern — Texas Instruments</i> <i>G. T. Zhou — Georgia Institute of Technology</i>	38
<b>Joint Estimation of Propagation Parameters in Multicarrier Systems</b> <i>S. Aouada and A. Belouchrani — Ecole Nationale Polytechnique, Algeria</i>	43
<b>OFDM Spectral Characterization: Estimation of the Bandwidth and the Number of Sub-Carriers</b> <i>W. Akmouche — CELAR, France</i> <i>E. Kerherve and A. Quinquis — ENSIETA, France</i>	48
<b>Blind Source Separation of Nonstationary Convolutively Mixed Signals</b> <i>B. S. Krongold and D. L. Jones — University of Illinois at Urbana-Champaign</i>	53
<b>A Versatile Spatio-Temporal Correlation Function for Mobile Fading Channels with Non-Isotropic Scattering</b> <i>A. Abdi and M. Kaveh — University of Minnesota</i>	58

## Session MA-2. Array Processing I

<b>A Batch Subspace ICA Algorithm</b> <i>A. Mansour and N. Ohnishi — RIKEN, Japan</i>	63
<b>Comparative Study of Two-Dimensional Maximum Likelihood and Interpolated Root-MUSIC with Application to Teleseismic Source Localization</b> <i>P.J. Chung and J. F. Böhme — Ruhr University, Germany</i> <i>A. B. Gershman — McMaster University, Canada</i>	68
<b>Bounds on Uncalibrated Array Signal Processing</b> <i>B. M. Sadler — Army Research Laboratory</i> <i>R. J. Kozick — Bucknell University</i>	73
<b>Array Processing in the Presence of Unknown Nonuniform Sensor Noise: A Maximum Likelihood Direction Finding Algorithm and Cramér-Rao Bounds</b> <i>M. Pesavento and A. B. Gershman — McMaster University, Canada</i>	78
<b>Matched Symmetrical Subspace Detector</b> <i>V. S. Golikov and F. C. Pareja — Ciencia y Tecnología del Mayab, A. C., Mexico</i>	83

# Table of Contents

<b>Multiple Source Direction Finding with an Array of M Sensors Using Two Receivers</b> <i>E. Fishler and H. Messer — Tel Aviv University, Israel</i> .....	86
<b>Self-Stabilized Minor Subspace Extraction Algorithm Based on Householder Transformation</b> <i>K. Abed-Meraim and S. Attallah — National University of Singapore, Singapore</i> <i>A. Chkeif — Telecom Paris, France</i> <i>Y. Hua — University of Melbourne, Australia</i> .....	90
<b>A Bootstrap Technique for Rank Estimation</b> <i>P. Pelin, R. Bricich and A. Zoubir — Curtin University of Technology, Australia.</i> .....	94
<b>Detection-Estimation of More Uncorrelated Sources than Sensors in NonInteger Sparse Linear Antenna Arrays</b> <i>Y. I. Abramovich and N. K. Spencer — CSSIP, Australia</i> .....	99
<b>A New Gerschgorin RadII Based Method for Source Number Detection</b> <i>H. Wu and C. Chen — Southern Taiwan University of Technology, Taiwan</i> .....	104

## Session MA-3. SPECTRUM ESTIMATION I

<b>Adapting Multitaper Spectrograms to Local Frequency Modulation</b> <i>J. W. Pitton — University of Washington</i> .....	108
<b>Optimal Subspace Selection for Non-Linear Parameter Estimation Applied to Refractivity from Clutter</b> <i>S. Kraut and J. Krolik — Duke University</i> .....	113
<b>MAP Model Order Selection Rule for 2-D Sinusoids in White Noise</b> <i>M. A. Kliger and J. M. Francos — Ben-Gurion University, Israel</i> .....	118
<b>Optimum Linear Periodically Time-Varying Filter</b> <i>D. Wei — Drexel University</i> .....	123
<b>Fast Approximated Sub-Space Algorithms</b> <i>M. A. Hasan — University of Minnesota Duluth</i> <i>A. A. Hasan — College of Electronic Engineering, Libya</i> .....	127
<b>Stochastic Algorithms for Marginal Map Retrieval of Sinusoids in Non-Gaussian Noise</b> <i>C. Andrieu and A. Doucet — University of Cambridge, UK</i> .....	131
<b>Harmonic Analysis Associated with Spatio-Temporal Transformations</b> <i>J. Leduc — Washington University in Saint Louis</i> .....	136

## Session MP-1. SIGNAL PROCESSING FOR COMMUNICATIONS II

<b>Blind Noise and Channel Estimation</b> <i>M. Frikel, W. Utschick, and J. Nossek — Technical University of Munich, Germany</i> .....	141
<b>Multiuser Detection in Impulsive Noise via Slowest Descent Search</b> <i>P. Spasojević — Rutgers University</i> <i>X. Wang — Texas A&amp;M University</i> .....	146
<b>Maximum Likelihood Delay-Doppler Imaging of Fading Mobile Communication Channels</b> <i>L. M. Davis — Bell Laboratories, Australia</i> <i>I. B. Collings — University of Sydney, Australia</i> <i>R. J. Evans — University of Melbourne, Australia</i> .....	151
<b>Enhanced Space-Time Capture Processing for Random Access Channels</b> <i>A. M. Kuzminskiy, K. Samaras, C. Luschi and P. Strauch — Bell Laboratories, Lucent Technologies, UK</i> .....	156
<b>Asymmetric Signaling Constellations for Phase Estimation</b> <i>T. Thaiupathump, C. D. Murphy and S. A. Kassam — University of Pennsylvania</i> .....	161
<b>A Convex Semi-Blind Cost Function for Equalization in Short Burst Communications</b> <i>K. K. Au and D. Hatzinakos — University of Toronto, Canada</i> .....	166

# Table of Contents

<b>Performance Analysis of Blind Carrier Phase Estimators for General QAM Constellations</b>	
<i>E. Serpedin — Texas A&amp;M University</i>	
<i>P. Ciblat and P. Loubaton — Université de Marne-la-Vallée, France</i>	
<i>G. B. Giannakis — University of Minnesota</i>	171
<b>Unbiased Parameter Estimation for the Identification of Bilinear Systems</b>	176
<i>S. Meddeb, J. Y. Tournet and F. Castanie — ENSEEIHT /TESA, France</i>	
<b>Blind Identification of Linear-Quadratic Channels with Usual Communication Inputs</b>	
<i>N. Petrochilos — Delft University of Technology, Netherlands</i>	
<i>P. Comon — Université de Nice, France</i>	181
<b>Joint Channel Estimation and Detection for Interference Cancellation in Multi-Channel Systems</b>	186
<i>C. Martin and B. Ottersten — Royal Institute of Technology (KTH), Sweden</i>	
<b>A Spatial Clustering Scheme for Downlink Beamforming in SDMA Mobile Radio</b>	191
<i>W. Huang and J. F. Doherty — Pennsylvania State University</i>	
<b>On the Use of Cyclostationary Filters to Transmit Information</b>	
<i>A. Duverdiér — CNES, France</i>	
<i>B. Lacaze and J. Tournet — ENSEEIHT/SIC, France</i>	196
<b>Non-Parametric Trellis Equalization in the Presence of Non-Gaussian Interference</b>	
<i>C. Luschi — Bell Laboratories, Lucent Technologies, UK</i>	
<i>B. Mulgrew — University of Edinburgh, UK</i>	201
<b>Analytical Blind Identification of a SISO Communication Channel</b>	206
<i>O. Grellier and P. Comon — Université de Nice, France</i>	
<b>The Role of Second-Order Statistics in Blind Equalization of Nonlinear Channels</b>	211
<i>R. López-Valcarce and S. Dasgupta — University of Iowa</i>	
<b>On Super-Exponential Algorithm, Constant Modulus Algorithm and Inverse Filter Criteria for Blind Equalization</b>	216
<i>C. Chi, C. Chen and B. Li — National Tsing Hua University, Taiwan</i>	
 <b>Session MP-2. STATISTICAL SIGNAL PROCESSING</b>	
<b>An Efficient Algorithm for Gaussian-Based Signal Decomposition</b>	221
<i>Z. Hong and B. Zheng — Xidian University, China</i>	
<b>Consistent Estimation of Signal Parameters in Non-Stationary Noise</b>	225
<i>J. Friedmann, E. Fishler and H. Messer — Tel Aviv University, Israel</i>	
<b>Channel Order and RMS Delay Spread Estimation for AC Power Line Communications</b>	
<i>H. Li — Stevens Institute of Technology</i>	
<i>Z. Bi and J. Li — University of Florida</i>	
<i>D. Liu — Watson Research Center</i>	
<i>P. Stoica — Uppsala University, Sweden</i>	229
<b>Taylor Series Adaptive Processing</b>	234
<i>D. J. Rabideau — Massachusetts Institute of Technology</i>	
<b>Adaptive Bayesian Signal Processing—A Sequential Monte Carlo Paradigm</b>	
<i>X. Wang and R. Chen — Texas A&amp;M University</i>	
<i>J. S. Liu — Stanford University</i>	239
<b>QQ-Plot Based Probability Density Function Estimation</b>	
<i>Z. Djurovic and V. Barroso — Instituto Superior Técnico — Instituto de Sistemas e Robótica, Portugal</i>	
<i>B. Kovacevic — University of Belgrade, Yugoslavia</i>	243
<b>Nonlinear System Inversion Applied to Random Variable Generation</b>	248
<i>A. Pagès-Zamora, M. A. Lagunas and X. Mestre — Universitat Politècnica de Catalunya, Spain</i>	
<b>The Numerical Spread as a Measure of Non-Stationarity: Boundary Effects in the Numerical Expected Ambiguity Function</b>	
<i>R. A. Hedges and B. W. Suter — Air Force Research Laboratory IFGC</i>	252

# Table of Contents

<b>Locally Stationary Processes</b>	
<i>M. E. Oxley and T. F. Reid — Air Force Institute of Technology</i>	
<i>B. W. Suter — Air Force Research Laboratory</i>	257
<b>Statistical Performance Comparison of a Parametric and a Non-Parametric Method for If Estimation of Random Amplitude Linear FM Signals in Additive Noise</b>	
<i>M. R. Morelande, B. Barkat and A. M. Zoubir — Curtin University of Technology, Australia</i>	262
<b>Session MP-3. RADAR SIGNAL PROCESSING</b>	
<b>The Application of a Nonlinear Inverse Noise Cancellation Technique to Maritime Surveillance Radar</b>	
<i>M. R. Cowper and B. Mulgrew — University of Edinburgh, UK</i>	267
<b>Adaptive Digital Beamforming RADAR for Monopulse Angle Estimation in Jamming</b>	
<i>K. Yu — GE Research &amp; Development Center</i>	
<i>D. J. Murrow — Lockheed Martin Ocean, Radar &amp; Sensors Systems</i>	272
<b>Statistical Analysis of SMF Algorithm for Polynomial Phase Signals Analysis</b>	
<i>A. Ferrari and G. Alengrin — Université de Nice Sophia-Antipolis, France</i>	276
<b>Passive Sonar Signature Estimation Using Bispectral Techniques</b>	
<i>R.K. Lennartsson, J.W.C. Robinson, and L. Persson — Defence Research Establishment, Sweden</i>	
<i>M.J. Hinich — University of Texas at Austin</i>	
<i>S. McLaughlin — University of Edinburgh, UK</i>	281
<b>Approximate CFAR Signal Detection in Strong Low Rank Non-Gaussian Interference</b>	
<i>I. P. Kirsteins — Naval Undersea Warfare Center</i>	
<i>M. Rangaswamy — ARCON Corporation</i>	286
<b>Blind Equalization of Phase Aberrations in Coherent Imaging: Medical Ultrasound and SAR</b>	
<i>S. D. Silverstein — University of Virginia</i>	291
<b>False Detection of Chaotic Behaviour in the Stochastic Compound K-Distribution Model of Radar Sea Clutter</b>	
<i>C.P. Unsworth, M.R. Cowper, S. McLaughlin, and B. Mulgrew — University of Edinburgh, UK</i>	296
<b>Session TA-1. BLIND SOURCE SEPARATION</b>	
<b>Recursive Estimator for Separation of Arbitrarily Kurtotic Sources</b>	
<i>M. Enescu and V. Koivunen — Helsinki Univ. of Technology, Finland</i>	301
<b>A Second Order Multi Output Deconvolution (SOMOD) Technique</b>	
<i>H. Bousbia-Salah and A. Belouchrani — Ecole Nationale Polytechnique, Algeria</i>	306
<b>DOA Estimation of Many W-Disjoint Orthogonal Sources from Two Mixtures Using Duet</b>	
<i>S. Rickard — Princeton University</i>	
<i>F. Dietrich — Siemens Corporate Research</i>	311
<b>Blind Separation of Non-Circular Sources</b>	
<i>J. Galy — LIRMM, France</i>	
<i>C. Adnet — Thomson-Csf Airsys, France</i>	315
<b>Blind Identification of Slightly Delayed Mixtures</b>	
<i>G. Chabriel and J. Barrère — Université de Toulon et du Var, France</i>	319
<b>Robust Source Separation Using Ranks</b>	
<i>L. Xiang, Y. Zhang and S. A. Kassam — University of Pennsylvania</i>	324
<b>Semi-Blind Maximum Likelihood Separation of Linear Convolutional Mixtures</b>	
<i>J. Xavier and V. Barroso — Instituto Superior Técnico — Instituto de Sistemas e Robótica, Portugal</i>	329
<b>Techniques for Blind Source Separation Using Higher-Order Statistics</b>	
<i>Z. M. Kamran and A. R. Leyman — Nanyang Technological University, Singapore</i>	
<i>K. Abed-Meraim — ENST/TSI, France</i>	334

## Table of Contents

<b>Joint-Diagonalization of Cumulant Tensors and Source Separation</b> <i>E. Moreau — MS-GESSY, ISITV, France</i>	339
<b>New Criteria for Blind Signal Separation</b> <i>N. Thirion-Moreau and E. Moreau — MS-GESSY, ISITV, France</i>	344
<b>An Iterative Algorithm Using Second Order Moments Applied to Blind Separation of Sources with Same Spectral Densities</b> <i>J. Cavassilas, B. Xerri and B. Borloz — Université de Toulon et du Var, France</i>	349
<b>Performance of Cumulant Based Inverse Filter Criteria for Blind Deconvolution of Multi-Input Multi-Output Linear Time-Invariant Systems</b> <i>C. Chi and C. Chen — National Tsing Hua University, Taiwan</i>	354
<b>Separation of Non Stationary Sources; Achievable Performance</b> <i>J. Cardoso — C.N.R.S./E.N.S.T., France</i>	359
<b>Modified BSS Algorithms Including Prior Statistical Information about Mixing Matrix</b> <i>J. Igual and L. Vergara — Universidad Politécnica Valencia, Spain</i>	364
<b>Approximate Maximum Likelihood Blind Source Separation with Arbitrary Source PDFs</b> <i>M. Ghogho and T. Durrani — University of Strathclyde, UK</i> <i>A. Swami — Army Research Lab</i>	368

### Session TA-2. SPECTRUM ESTIMATION II

<b>Power Spectral Density Analysis of Randomly Switched Pulse Width Modulation for DC/AC Converters</b> <i>R. L. Kirlin — University of Victoria, Canada</i> <i>M. M. Bech — University of Aalborg, Denmark</i> <i>A.M. Trzynadlowski — University of Nevada Reno</i>	373
<b>Study on Spectral Analysis and Design for DC/DC Conversion Using Random Switching Rate PWM</b> <i>R. L. Kirlin, J. Wang, and R. M. Dizaji — University of Victoria, Canada</i>	378
<b>Spectral Subtraction and Spectral Estimation</b> <i>M. A. Lagunas and A. I. Perez-Neira — Campus Nord UPC, Spain</i>	383
<b>Parameter Estimation: The Ambiguity Problem</b> <i>V. Lefkaditis and A. Manikas — Imperial College of Science, Technology and Medicine, UK</i>	387
<b>On Multiwindow Estimators for Correlation</b> <i>A. Hanssen — University of Tromsø, Norway</i>	391
<b>Asymptotic Analysis of the Least Squares Estimate of 2-D Exponentials in Colored Noise</b> <i>G. Cohen and J. M. Francos — Ben-Gurion University, Israel</i>	396
<b>Cross-Spectral Methods for Processing Biological Signals</b> <i>D. J. Nelson — Department of Defense</i>	400
<b>Default Prior for Robust Bayesian Model Selection of Sinusoids in Gaussian Noise</b> <i>C. Andrieu — Cambridge University, UK</i> <i>J.-M. Pérez — Universidad Simón Bolívar, Venezuela</i>	405
<b>On the Exact Solution to the "Gliding Tone" Problem</b> <i>L. Galleani and L. Cohen — City University of New York</i>	410
<b>Baseline and Distribution Estimates of Complicated Spectra</b> <i>D. J. Thomson — Bell Labs</i>	414

### Session TA-3. ARRAY PROCESSING II

<b>Distributed Source Localization with Multiple Sensor Arrays and Frequency-Selective Spatial Coherence</b> <i>R. J. Kozick — Bucknell University</i> <i>B. M. Sadler — Army Research Laboratory</i>	419
---	-----



## Table of Contents

<b>Deterministic Maximum Likelihood DOA Estimation in Heterogeneous Propagation Media</b>	
<i>P. Stoica — Uppsala University, Sweden</i>	
<i>O. Besson — ENSICA, France</i>	
<i>A. B. Gershman — McMaster University, Canada</i> .....	424
<b>Efficient Signal Detection in Perturbed Arrays</b>	
<i>A. M. Rao and D. L. Jones — University of Illinois</i> .....	429
<b>A Neural Network Approach for DOA Estimation and Tracking</b>	
<i>L. Badidi and L. Radouane — LESSI, Morocco</i> .....	434
<b>Partially Adaptive Array Algorithm Combined with CFAR Technique in Transform Domain</b>	
<i>S. Moon, D. Yun, and D. Han — Kyungpook National University, Korea</i> .....	439
<b>A New Beamforming Algorithm Based on Signal Subspace Eigenvectors</b>	
<i>M. Biguesh and M. H. Bastani — Sharif University of Technology, Iran</i>	
<i>S. Valaee — Tarbiat Modares University, Iran</i>	
<i>B. Champagne — McGill University, Canada</i> .....	444
<b>Detection of Sources in Array Processing Using the Bootstrap</b>	
<i>R. Brich, P. Pelin and A. Zoubir — Curtin University of Technology, Australia</i> .....	448
<b>Robust Localization of Scattered Sources</b>	
<i>J. Tabrikian — Ben-Gurion University, Israel</i>	
<i>H. Messer — Tel Aviv University, Israel</i> .....	453
 <b>Session TP-1. APPLICATION OF JOINT TIME-FREQUENCY TECHNIQUES IN RADAR PROCESSING</b>	
<b>ISAR Imaging and Crystal Structure Determination from EXAFS Data Using a Super-Resolution Fast Fourier Transform</b>	
<i>G. Zweig — Signition, Inc.</i>	
<i>B. Wohlberg — Los Alamos National Laboratory</i> .....	458
<b>Analysis of Radar Micro-Doppler Signature With Time-Frequency Transform</b>	
<i>V. C. Chen — Naval Research Laboratory</i> .....	463
<b>Estimating the Parameters of Multiple Wideband Chirp Signals in Sensor Arrays</b>	
<i>A. B. Gershman and M. Pesavento — McMaster University, Canada</i>	
<i>M. G. Amin — Villanova University</i> .....	467
<b>On the Use of Space-Time Adaptive Processing and Time-Frequency Data Representations for Detection of Near-Stationary Targets in Monostatic Clutter</b>	
<i>D. C. Braunreiter, H.-W. Chen, M. L. Cassabaum, J. G. Riddle, A. A. Samuel, J. F. Scholl and H. A. Schmitt — Raytheon Missile Systems</i> .....	472
<b>Application of Adaptive Joint Time-Frequency Processing to ISAR Image Formation</b>	
<i>H. Ling and J. Li — University of Texas at Austin</i> .....	476
<b>Joint Time-Frequency Analysis of SAR Data</b>	
<i>R. Fiedler and R. Jansen — Naval Research Laboratory</i> .....	480
<b>Pulse Propagation in Dispersive Media</b>	
<i>L. Cohen — City University of New York</i> .....	485
 <b>Session TP-2. NETWORK TRAFFIC MODELING</b>	
<b>Wavelet-Based Models for Network Traffic</b>	
<i>D. Wei and H. Cheng — Drexel University</i> .....	490
<b>The Extended On/Off Process for Modeling Traffic in High-Speed Communication Networks</b>	
<i>X. Yang, A. P. Petropulu and V. Adams — Drexel University</i> .....	495
<b>A Simulation Study of the Impact of Switching Systems on Self-Similar Properties of Traffic</b>	
<i>Y. Zhou and H. Sethu — Drexel University</i> .....	500

# Table of Contents

<b>Parameter Estimation in Farima Processes with Applications to Network Traffic Modeling</b> <i>J. Ilow — Dalhousie University, Canada</i>	505
 <b>Session TP-3. SIGNAL PROCESSING FOR GPS</b>	
<b>Nonlinear Filtering Algorithm with Its application In INS Alignment</b> <i>R. Zhao and Q. Gu — Tsinghua University, China</i>	510
<b>GPS Jammer Suppression with Low-Sample Support Using Reduced-Rank Power Minimization</b> <i>W. L. Myrick and M. D. Zoltowski — Purdue University</i> <i>J. S. Goldstein — SAIC</i>	514
<b>Jammer Excision in Spread Spectrum Using Discrete Evolutionary-Hough Transform and Singular Value Decomposition</b> <i>R. Suleesathira and L. F. Chaparro — University of Pittsburgh</i>	519
<b>Spatial and Temporal Processing of GPS Signals</b> <i>P. Xiong and S. N. Batalama — State University of New York at Buffalo</i> <i>M. J. Medley — Air Force Research Laboratory</i>	524
<b>Subspace Projection Techniques for Anti-FM Jamming GPS Receivers</b> <i>L. Zhao and M. G. Amin — Villanova University</i> <i>A. R. Lindsey — Air Force Research Laboratory</i>	529
 <b>Session TP-4. WAVELETS</b>	
<b>Fixed-Point HAAR-Wavelet-Based Echo Cancellor</b> <i>M. Doroslovački and I. Khan — George Washington University</i> <i>B. Kosanović — Texas Instruments</i>	534
<b>Wavelet-Polyspectra: Analysis of Non-Stationary and Non-Gaussian/Non-Linear Signals</b> <i>Y. Larsen and A. Hanssen — University of Tromsø, Norway</i>	539
<b>Adaptive Seismic Compression by Wavelet Shrinkage</b> <i>M.F. Khène and S.H. Abdul-Jauwad — King Fahd University of Petroleum &amp; Minerals, Saudi Arabia</i>	544
<b>Representations of Stochastic Processes Using COIFLET-Type Wavelets</b> <i>D. Wei and H. Cheng — Drexel University</i>	549
 <b>Session WA-1. TIME-FREQUENCY ANALYSIS</b>	
<b>Time-Frequency Coherence Analysis of Nonstationary Random Processes</b> <i>G. Matz and F. Hlawatsch — Vienna University of Technology, Austria</i>	554
<b>Multi-Component IF Estimation</b> <i>Z. M. Hussain and B. Boashash — Queensland University of Technology, Australia</i>	559
<b>Detection of Seizures in Newborns Using Time-Frequency Analysis of EEG Signals</b> <i>B. Boashash, H. Carson and M. Mesbah — Queensland University of Technology, Australia</i>	564
<b>Multitaper Reduced Interference Distribution</b> <i>S. Aviyente and W. J. Williams — University of Michigan</i>	569
<b>Instantaneous Spectral Skew and Kurtosis</b> <i>P. J. Loughlin and K. L. Davidson — University of Pittsburgh</i>	574
<b>Adaptive Time-Frequency Representations for Multiple Structures</b> <i>A. Papandreou-Suppappola — Arizona State University</i> <i>S. B. Suppappola — Pipeline Technologies, Inc.</i>	579
<b>A Resolution Performance Measure for Quadratic Time-Frequency Distributions</b> <i>B. Boashash and V. Sucic — Queensland University of Technology, Australia</i>	584
<b>The Wigner Distribution for Ordinary Linear Differential Equations and Wave Equations</b> <i>L. Galleani and L. Cohen — City University of New York</i>	589

# Table of Contents

<b>Application of Time-Frequency Techniques for the Detection of Anti-Personnel Landmines</b> <i>B. Barkat, A.M. Zoubir and C.L. Brown — Curtin University of Technology, Australia</i>	594
<b>A New Matrix Decomposition Based on Optimum Transformation of the Singular Value Decomposition Basis Sets Yields Principal Features of Time-Frequency Distributions</b> <i>D. Groutage — Naval Surface Warfare Center</i> <i>D. Bennink — Applied Measurements Systems Intl.</i>	598
<b>Minimum Entropy Time-Frequency Distributions</b> <i>A. El-Jaroudi — University of Pittsburgh</i>	603
<b>Uncertainty in the Time-Frequency Plane</b> <i>P. M. Oliveira — Escola Naval, Portugal</i> <i>V. Barroso — Instituto Superior Técnico ISR/DEEC, Portugal</i>	607
<b>High Resolution Frequency Tracking via Non-Negative Time-Frequency Distributions</b> <i>R. M. Nickel and W. J. Williams — University of Michigan</i>	612
 <b>Session WA-2. HIGHER-ORDER SPECTRAL ANALYSIS</b>	
<b>A Cumulant Subspace Approach to FIR Multuser Channel Estimation</b> <i>J. Liang and Z. Ding — University of Iowa</i>	616
<b>An Efficient Forth Order System Identification (FOSI) Algorithm Utilizing the Joint Diagonalization Procedure</b> <i>A. Belouchrani — Ecole National Polytechnique, Algeria</i> <i>B. Derras — Cirrus Logic Inc.</i>	621
<b>Unity-Gain Cumulant-Based Adaptive Line Enhancer</b> <i>R. R. Gharieb and A. Cichocki — RIKEN, Japan</i> <i>Y. Horita and T. Murai — Toyama University, Japan</i>	626
<b>Adaptive Detection and Extraction of Sparse Signals Embedded in Colored Gaussian Noise Using Higher Order Statistics</b> <i>R. R. Gharieb and A. Cichocki — RIKEN, Japan</i> <i>S. F. Filipowicz — Warsaw University of Technology, Poland</i>	631
<b>Higher-Order Matched Field Processing</b> <i>R. M. Dizaji, R. L. Kirlin, and N. R. Chapman — University of Victoria, Canada</i>	635
<b>Multwindow Bispectral Estimation</b> <i>Y. Birkelund and A. Hanssen — University of Tromsø, Norway</i>	640
 <b>WA-3. SIGNAL PROCESSING FOR COMMUNICATIONS III</b>	
<b>Global Convergence of a Single-Axis Constant Modulus Algorithm</b> <i>A. Shah, S. Biracree, R. A. Casas, T. J. Endres, S. Hulyalkar, T. A. Schaffer, and C. H. Strolle — NxtWave Communications</i>	645
<b>A Novel Modulation Method for Secure Digital Communications</b> <i>A. Salberg and A. Hanssen — University of Tromsø, Norway</i>	650
<b>A Multitime-Frequency Approach for Detection and Classification of Noisy Frequency Modulations</b> <i>M. Colas, G. Gelle, and G. Delaunay — L.A.M.-URCA, France</i> <i>J. Galy — L.I.R.M.M., France</i>	655
<b>NDA PLL Design for Carrier Phase Recovery of QPSK/TDMA Bursts without Preamble</b> <i>J. Lee — COMSAT Laboratories</i>	660
<b>An Optimized Multi-Tone Calibration Signal for Quadrature Receiver Communication Systems</b> <i>R. A. Green — North Dakota State University</i>	664
<b>A Polynomial Rooting Approach for Synchronization in Multipath Channels Using Antenna Arrays</b> <i>G. Seco and J.A. Fernández-Rubio — Univ. Politècnica de Catalunya, Spain</i> <i>A. L. Swindlehurst — Brigham Young University</i>	668

## Table of Contents

<b>Super-Exponential-Estimator for Fast Blind Channel Identification of Mobile Radio Fading Channels</b>	
<i>A. Schmidbauer — Munich University of Technology, Germany</i>	673
<b>Finite Data Record Maximum SINR Adaptive Space-Time Processing</b>	
<i>I. N. Psaromiligkos and S. N. Batalama — State University of New York at Buffalo</i>	677
<b>On the Effects of Rotating Blades on DS/SS Communication Systems</b>	
<i>Y. Zhang and M. G. Amin — Villanova University</i>	
<i>V. Mancuso — Boeing Helicopter Division</i>	682
<b>Joint Synchronization and Symbol Detection In Asynchronous DS-CDMA Systems</b>	
<i>F. Rey, G. Vázquez, and J. Riba — Polytechnic University of Catalonia, Spain</i>	687
<b>New Criteria for Blind Equalization of M-PSK Signals</b>	
<i>Z. Xu and P. Liu — University of California</i>	692
<b>Third-Order Blind Equalization Properties of Hexagonal Constellations</b>	
<i>C. D. Murphy — Helsinki University of Technology, Finland</i>	697
 <b>Session WA-4. ACOUSTICAL SIGNAL PROCESSING</b>	
<b>Comparison of the Cyclostationary and the Bilinear Approaches: Theoretical Aspects and Applications to Industrial Signals.</b>	
<i>L. Bouillaut and M. Sidahmed — Université de Technologie de Compiègne, France</i>	702
<b>Array Processing of Underwater Acoustic Sensors Using Weighted Fourier Integral Method</b>	
<i>I. S. D. Solomon and A. J. Knight — Defence Science and Technology Organisation, Australia</i>	707
<b>A Hierarchical Algorithm for Nearfield Acoustic Imaging</b>	
<i>M. Peake and M. Karan — CSSIP, Australia</i>	
<i>D. Gray — University of Adelaide, Australia</i>	712
<b>An Introduction to Synthetic Aperture Sonar</b>	
<i>D. Marx, M. Nelson, E. Chang, W. Gillespie, A. Putney, and K. Warman — Dynamics Technology, Inc.</i>	717
<b>Classification of Acoustic and Seismic Data Using Nonlinear Dynamical Signal Models</b>	
<i>R. K. Lennartsson — Defence Research Establishment, Sweden</i>	
<i>Á. Péntek and J. B. Kadtké — University of California</i>	722
<b>The Performance of Sparse Time-Reversal Mirrors In the Context of Underwater Communications</b>	
<i>J. Gomes and V. Barroso — Instituto Superior Técnico — Instituto de Sistemas e Robótica, Portugal</i>	727
<b>Beam Patterns of an Underwater Acoustic Vector Hydrophone</b>	
<i>K. T. Wong — Chinese University of Hong Kong, China</i>	
<i>H. Chi — Purdue University</i>	732

# MULTISTAGE MULTIUSER DETECTION FOR CDMA WITH SPACE-TIME CODING

Yumin Zhang and Rick S. Blum

EECS Department, Lehigh University  
Bethlehem, PA 18015  
rblum@eecs.lehigh.edu

## ABSTRACT

*The combination of Turbo codes and space-time block codes is studied for use in CDMA systems. Each user's data are first encoded by a Turbo code. The Turbo coded data are next sent to a space-time block encoder which employs a BPSK constellation. The space-time encoder output symbols are transmitted through the fading channel using multiple antennas. A multistage receiver is proposed using non-linear MMSE estimation and a parallel interference cancellation scheme. Simulations show that with reasonable levels of multiple access interference ( $\rho \leq 0.3$ ), near single user performance is achieved. The receiver structure is generalized to decode CDMA signals with space-time convolutional coding and similar performance is observed.*

## 1. INTRODUCTION

Space-time codes [1]-[4] use multiple transmit and receive antennas to achieve diversity and coding gain for communication over fading channels. High bandwidth efficiency is achieved, with performance close to the theoretical outage capacity [1]. Turbo codes [5] are a family of powerful channel codes, which have been shown to achieve near Shannon capacity over additive white Gaussian noise channels. Since their introduction, both space-time codes and Turbo codes have received considerable attention. In the CDMA2000 Radio Transmission Technology (RTT) proposed for the third generation systems, both space-time codes and Turbo codes have been adopted [6].

Although papers treating either just space-time codes or Turbo codes abound, jointly considering space-time codes and Turbo codes in CDMA systems is a relatively new topic. In this paper, we initiate a study on this topic where we focus on space-time block codes [3][4]. Our research develops suboptimum low-complexity receivers, which will be needed.

This paper is organized as follows. Section 2 first sets up the system configuration and develops the received signal model. A brief review of space-time block

codes is given in Section 3. The structure of our multistage receiver is discussed in Section 4. Section 5 presents simulation results. Conclusions are given in Section 6.

## 2. SYSTEM CONFIGURATION AND RECEIVED SIGNAL MODEL

Fig. 2 depicts a  $K$  user synchronous CDMA system with combined Turbo coding and space-time block coding. There are  $N$  transmit antennas and  $M$  receive antennas in the system. Suppose user  $k$ ,  $k = 1, \dots, K$ , has a block of binary information bits  $\{d_k(i), i = 1, \dots, L_1\}$  to transmit. These bits are first encoded by a Turbo code with rate  $R_1 = \frac{L_1}{L_2}$ . The bits which are produced by the Turbo encoder, denoted by  $\{\tilde{d}_k(i), i = 1, \dots, L_2\}$ , are passed to a space-time block encoder. This space-time block code uses a transmission matrix  $\mathcal{G}_N$  [3] with a BPSK constellation, generates  $N$  output bits during each time slot, and has rate  $R_2 = \frac{L_2}{L}$ . During time slot  $l$ ,  $N$  bits are transmitted, which are denoted by  $\{b_{nk}(l), n = 1, \dots, N\}$ , for  $l = 1, \dots, L$ . The bit  $b_{nk}(l) \in \{-1, +1\}$  is spread using a unique spreading waveform  $s_k(t)$  and transmitted using antenna  $n$ . For convenience we denote the vector of  $n$ th output bits from all  $K$  users as  $\mathbf{b}_n(l) = [b_{n1}(l), \dots, b_{nK}(l)]^T$ , and we note that all of these bits are transmitted by antenna  $n$  during time slot  $l$ . We define the set of bits  $\{\mathbf{b}_n(l), l = 0, \dots, L-1\}$  as one frame of data.

The fading coefficient for the path between transmit antenna  $n$  and receive antenna  $m$  is denoted by  $\alpha_{nm}$ . In our research, we assume a flat quasi-static fading environment [3], where the fading coefficients are constant during a frame and are independent from one frame to another. Further we assume for simplicity that perfect estimates of all fading coefficients are available at the receiver. The received signal at antenna  $m$  is

$$r_m(t) = \sum_{n=1}^N \sum_{k=1}^K \sum_{l=0}^{L-1} \alpha_{nm} A_k b_{nk}(l) s_k(t-lT) + \eta_m(t) \quad (1)$$

where  $T$  is the bit period,  $A_k$  is the transmitted signal

amplitude for user  $k$ , and  $\eta_m(t)$  is the complex channel noise at receive antenna  $m$ . The received signal  $r_m(t)$  is next passed through a matched filter bank, with each filter matched to one user's spreading waveform. Denote the matched filter outputs at receive antenna  $m$  for the time slot  $j$  by  $\mathbf{y}_m(j) = [y_{m1}(j), \dots, y_{mK}(j)]^T$ . The equation describing  $\mathbf{y}_m(j)$  can be represented in vector form as

$$\mathbf{y}_m(j) = \mathbf{R}\mathbf{A} \sum_{n=1}^N \alpha_{nm} \mathbf{b}_n(j) + \mathbf{n}_m(j) \\ m = 1, \dots, M, j = 0, \dots, L-1. \quad (2)$$

where  $\mathbf{R}$  is the  $K \times K$  cross-correlation matrix of the spreading codes,  $\mathbf{A} = \text{diag}(A_1, \dots, A_K)$ , and  $\mathbf{n}_m(j)$  is the  $K \times 1$  complex noise vector after matched filtering. Assuming the channel noise is Gaussian with zero mean and autocorrelation function  $\sigma^2 \delta(\tau)$ ,  $\mathbf{n}_m(j)$  has a multidimensional Gaussian distribution  $N(0, \sigma^2 \mathbf{R})$ .

### 3. SPACE-TIME BLOCK CODES

An extensive discussion of space-time block codes is given in [3][4]. Here we consider only  $N = 2$  antenna cases. Extension to  $N > 2$  cases is straightforward. A BPSK space-time block code with two transmit antennas is described by the transmission matrix

$$\mathcal{G}_2 = \begin{pmatrix} s_1 & s_2 \\ -s_2 & s_1 \end{pmatrix}. \quad (3)$$

The encoder works as follows. The block of  $L_2$  Turbo coded bits enter the encoder and are grouped into units of two bits. Each group of two bits are mapped to a pair of BPSK symbols  $s_1$  and  $s_2$ . These symbols are transmitted during two consecutive time slots. During the first time slot,  $s_1$  and  $s_2$  are transmitted simultaneously from antenna one and two respectively. During the second time slot,  $-s_2$  and  $s_1$  are transmitted simultaneously from antenna one and two, respectively. The code rate of  $\mathcal{G}_2$  is 1.

In [3][4], the transmission matrix is designed so that the columns are orthogonal to each other. This allows a simple receiver structure using only linear processing. We illustrate this using the code described in (3) as an example. Extension to  $N > 2$  cases is straightforward. Assuming there are  $M$  receive antennas, the received signal at antenna  $m$  during the first and second time slots, denoted by  $y_m(1)$  and  $y_m(2)$ , are

$$y_m(1) = \alpha_{1m} s_1 + \alpha_{2m} s_2 + n_m(1) \\ y_m(2) = -\alpha_{1m} s_2 + \alpha_{2m} s_1 + n_m(2) \quad (4)$$

where  $n_m(1)$  and  $n_m(2)$  are two iid complex Gaussian noise samples with variance  $\sigma_n^2$ . The observations in

(4) can be combined to yield the improved quantities  $\tilde{s}_1$  and  $\tilde{s}_2$  using

$$\begin{aligned} \tilde{s}_1 &= \alpha_{1m}^* y_m(1) + \alpha_{2m} y_m^*(2) \\ &= (|\alpha_{1m}|^2 + |\alpha_{2m}|^2) s_1 + \alpha_{1m}^* n_m(1) + \alpha_{2m} n_m^*(2) \\ \tilde{s}_2 &= \alpha_{2m}^* y_m(1) - \alpha_{1m} y_m^*(2) \\ &= (|\alpha_{1m}|^2 + |\alpha_{2m}|^2) s_2 + \alpha_{2m}^* n_m(1) - \alpha_{1m} n_m^*(2) \end{aligned}$$

Combining quantities obtained at each receive antenna yields

$$\begin{aligned} \tilde{\tilde{s}}_1 &= \sum_{m=1}^M (\alpha_{1m}^* y_m(1) + \alpha_{2m} y_m^*(2)) \equiv C s_1 + n_1 \\ \tilde{\tilde{s}}_2 &= \sum_{m=1}^M (\alpha_{2m}^* y_m(1) - \alpha_{1m} y_m^*(2)) \equiv C s_2 + n_2 \end{aligned} \quad (5)$$

where

$$C = \sum_{m=1}^M (|\alpha_{1m}|^2 + |\alpha_{2m}|^2). \quad (6)$$

The Gaussian noise variables  $n_1$  and  $n_2$  have variance

$$\sigma_b^2 = \sigma_n^2 \sum_{m=1}^M (|\alpha_{1m}|^2 + |\alpha_{2m}|^2) \quad (7)$$

It is easily seen from (5), (6) and (7) that after this simple linear combining, the resulting signals are equivalent to those obtained from using maximal ratio combining [7] techniques for systems with 1 transmit antenna and  $2M$  receive antennas. This combining technique will be used in two places in our low-complexity receiver as discussed in the next section.

### 4. LOW-COMPLEXITY MULTISTAGE RECEIVER

The optimum receiver that minimizes the frame error rate should construct a "super-trellis" for decoding. The super-trellis combines the trellis of Turbo codes and the structure of the multiuser channel and space-time block codes. Due to the interleavers used in the Turbo codes, it is very hard to construct such a super-trellis. In fact, "optimum decoding" for Turbo codes alone is impossible in practice. This is why suboptimum iterative decoding schemes are used to decode Turbo codes [5]. Thus instead of trying to find an optimum receiver, which would obviously have a prohibitively high complexity, our goal in this section is to develop a low-complexity suboptimum receiver.

We suggest the multistage receiver structure depicted in Fig. 2. The output of the matched filter bank is first passed to a decorrelating detector [8], which attempts to eliminate the multiple access interference (MAI) completely with perfect estimation. The output

of the decorrelating detector at receive antenna  $m$  and time slot  $j$  is

$$\tilde{\mathbf{y}}_m(j) = (\mathbf{R}\mathbf{A})^{-1}\mathbf{y}_m(j) = \sum_{n=1}^N \alpha_{nm}\mathbf{b}_n(j) + \tilde{\mathbf{n}}_m(j) \quad (8)$$

where we defined the noise vector  $\tilde{\mathbf{n}}_m(j) = (\mathbf{R}\mathbf{A})^{-1}\mathbf{n}_m(j)$ , which has a Gaussian distribution with covariance matrix

$$\tilde{\mathbf{R}} = \sigma^2(\mathbf{A}\mathbf{R}\mathbf{A})^{-1}. \quad (9)$$

The elements from  $\tilde{\mathbf{y}}_1(j)$ , ...,  $\tilde{\mathbf{y}}_M(j)$  corresponding to the  $k$ th user, denoted by  $\tilde{y}_{1k}(j), \dots, \tilde{y}_{Mk}(j)$ , are combined using the technique discussed in Section 3 to provide improved observations for user  $k$ . These improved observations are sent to a single user Turbo decoder to perform the first stage of decoding. The Turbo decoder produces posterior probabilities for user  $k$ 's transmitted bits. These posterior probabilities, together with the diversity combined observations, are used by a soft estimator to form soft estimates of user  $k$ 's transmitted bits.

The soft estimator uses non-linear minimum mean square error (MMSE) estimation [9] to form the soft estimates. From (5), it is seen that the diversity combined observations for user  $k$  can always be represented in the form of  $y = Cb + n$ , where  $y$  is the noisy observation,  $b$  is the transmitted bit,  $C$  is a known constant and  $n$  is a complex Gaussian noise sample with variance denoted by  $\sigma_b^2$ . The soft estimate of  $b$  is obtained by

$$E\{b|y\} = \frac{\frac{Pr(b=+1)}{Pr(b=-1)} e^{\frac{2Re(Cy^*)}{\sigma_b^2}} - e^{-\frac{2Re(Cy^*)}{\sigma_b^2}}}{\frac{Pr(b=+1)}{Pr(b=-1)} e^{\frac{2Re(Cy^*)}{\sigma_b^2}} + e^{-\frac{2Re(Cy^*)}{\sigma_b^2}}}, \quad (10)$$

where the prior probabilities  $Pr(b = \pm 1)$  can be updated using the posterior probabilities obtained by the Turbo decoders.

The transmitted signals are reconstructed using the soft estimates as if they were binary digits. Denote the reconstructed encoder output for antenna  $n$  and user  $k$  during time slot  $j$  as  $\hat{b}_{nk}(j)$  and define  $\hat{\mathbf{b}}_n(j) = [\hat{b}_{n1}(j), \dots, \hat{b}_{nK}(j)]^T$ . The reconstructed signals  $\{\hat{\mathbf{b}}_n(j), n = 1, \dots, N, j = 0, \dots, L-1\}$  are used in soft MAI cancellation to produce "cleaner" received signals for each user. To cancel MAI for user  $k$ , we first define a vector  $\hat{\mathbf{b}}_n^{(k)}(j)$  equal to  $\hat{\mathbf{b}}_n(j)$  except that its  $k$ th element is zero. The MAI-reduced observation for user  $k$  at receive antenna  $m$  is obtained using

$$\mathbf{y}_m^{(k)}(j) = \mathbf{y}_m(j) - \mathbf{R}\mathbf{A} \sum_{n=1}^N \alpha_{nm}\hat{\mathbf{b}}_n^{(k)}(j) \quad (11)$$

When perfect estimate of  $\mathbf{b}_n(j)$  is available,  $\mathbf{y}_m^{(k)}(j)$  offers  $K$  different observations of the signal from user  $k$ , contaminated only by channel noise. For simplicity, we use the  $k$ th element of  $\mathbf{y}_m^{(k)}(j)$  for processing, which gives the highest SNR for user  $k$ . The  $k$ th elements of  $\mathbf{y}_m^{(k)}(j)$ ,  $m = 1, \dots, M$ , at all receive antennas are combined using the techniques discussed in Section 3. The improved observations are passed to another set of Turbo decoders to perform the second stage of decoding. These Turbo decoders produce the final "hard" decisions on each user's transmitted bits.

## 5. SIMULATION RESULTS

Monte Carlo simulations are carried out to study the performance of the proposed multistage receiver. Consider a 4 user synchronous CDMA system with 2 transmit antennas and 2 receive antennas. Each user's bits are first encoded by a rate 1/3 Turbo code with constraint length  $\nu = 5$  and generator 23, 35 (octal form). The random interleaver chosen for the Turbo code has length 128. The block of Turbo coded data is encoded using a space-time block code with the code matrix  $\mathcal{G}_2$  from (3) and a BPSK constellation. Next the output bits are spread using each user's spreading waveform and the results are transmitted using 2 antennas over the fading channel. The path gains are modeled as samples of independent complex Gaussian random variables with variance 0.5 per dimension (real or imaginary). Quasi-static fading is assumed. For the CDMA channel, we use the symmetric channel model where the cross-correlation between all pairs of two users is the common value  $\rho$ . The SNR for user  $k$  is defined as

$$SNR_k = \frac{NA_k}{\sigma^2 R_1 R_2} \quad (12)$$

Fig. 3 gives the BER performance of the proposed multistage receiver in Gaussian noise when all users have the same power ( $\mathbf{A} = \mathbf{I}$ ). The BER performance for the first stage and second stage decoding are both plotted, which we denote by "S1" and "S2" on the graph. For comparison, we also give the single user performance, which is the Turbo code performance for the fading channel under consideration. The performance of the space-time block code using  $\mathcal{G}_2$  without the Turbo coding is also shown. For  $\rho = 0.1$ , single user performance is nearly achieved after just the first stage decoding. The second stage decoding curve is indistinguishable from that of the single user performance. For  $\rho = 0.3$ , the performance improvement obtained by employing the second stage of decoding is obvious from Fig. 3b. After the second stage decoding, single user performance is approached. By combining a Turbo code with a space-time block code, a performance gain

of about 2.5dB is achieved at  $\text{BER}=10^{-4}$  compared to using a space-time block code only.

An iterative receiver structure can be easily constructed by feeding back the posterior information obtained after the second stage decoding to the soft estimators. We have carried out simulations using this iterative structure, but results show that the improvement over the second stage of decoding is marginal. In Fig. 3b, we plot the BER performance for the second iteration of the "iterative receiver" (denoted by "Ite 2"), which is almost indistinguishable from the second stage decoding curve. Thus the extra computations incurred by the iterative structure are not justified.

Next we study the performance of our receiver in a near-far situation where two users are 20dB stronger than the other two users, all other parameters remain the same as in Fig 3. The BER performance for the strong user and weak user are given in Fig. 4a and 4b respectively. The performance, for both the weak and strong users, approaches single user performance after the second stage decoding.

Finally, we point out that the received signal model in (2) is also valid for a CDMA system with space-time convolutional coding [1] replacing the combination of space-time block codes and Turbo codes. An iterative receiver can be constructed using the parallel interference cancellation scheme [10]. Fig. 1 gives the frame error rate performance for the first two iterations of the iterative receiver for a CDMA system with space-time convolutional coding. It is seen that with 2 iterations, single user performance is achieved. Another observation is that the performance improvement obtained by employing the iterative structure is marginal. This is consistent with our previous observations for the space-time block coded system.

## 6. CONCLUSIONS

In this paper, we studied the application of Turbo codes and space-time block codes in CDMA systems. A multistage receiver is proposed using parallel interference cancellation schemes. Simulation results show that with reasonable levels of MAI ( $\rho \leq 0.3$ ), near single user performance can be achieved. The receiver developed in this paper was generalized to decode CDMA signals with space-time convolutional coding and similar performance was observed.

## 7. REFERENCES

[1] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: Performance criteria and code construction," *IEEE Trans. Info. Theo.*, vol. 44, No. 2, pp. 744-765, Mar. 1998.

[2] S. M. Alamouti, "A simple transmitter diversity scheme for wireless communications," *IEEE JSAC*, vol. 16, No. 8, pp. 1451-1458, Oct. 1998.

[3] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-time block coding for wireless communications: performance results," *IEEE JSAC*, vol. 17, No. 3, pp. 451-460, March 1999.

[4] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-time block codes from orthogonal designs," *IEEE Trans. Info. Theo.*, vol. 45, No. 5, pp. 1456-1467, July 1999.

[5] C. Berrou and A. Glavieux, "Near optimum error correcting coding and decoding: Turbo-Codes," *IEEE Trans. Comm.*, vol. 44, No. 10, pp. 1261-1271, Oct. 1996.

[6] S. Dennett, "The CDMA2000 ITU-R RTT candidate submission," V. 0.17, TIA, July 28, 1998.

[7] J. G. Proakis, *Digital Communications*, 3rd Edition, McGraw-Hill, 1995.

[8] S. Verdú, *Multuser Detection*, UK: Cambridge University Press, 1998.

[9] A. Papoulis, *Probability, Random Variables and Stochastic Processes*, New York: McGraw-Hill, 1984.

[10] Yumin Zhang, *Iterative and Adaptive Receivers For Wireless Communication and Radar Systems*, Ph.D. Dissertation, Lehigh University, May 2000.

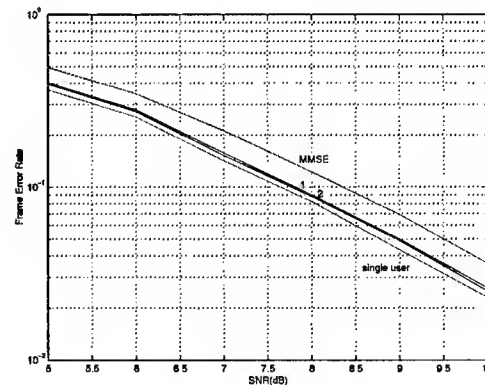


Figure 1: Performance of the iterative multiuser receiver for CDMA with space-time convolutional coding [10] with  $K = 4$ ,  $\rho = 0.3$ , 4-PSK S-T code with rate  $2/b/s/Hz$ , 130 symbols per frame, 2 transmit and 2 receive antennas where MMSE is used in the first stage decoding.



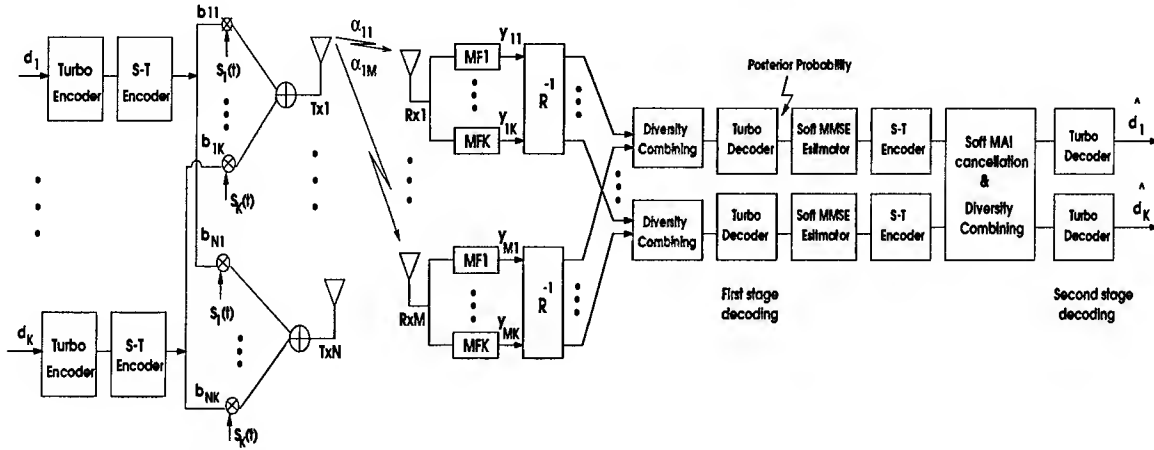


Figure 2: Structure of our  $K$  user CDMA system (including our multistage receiver) with combined Turbo coding and space-time block coding,  $N$  transmit antennas and  $M$  receive antennas.

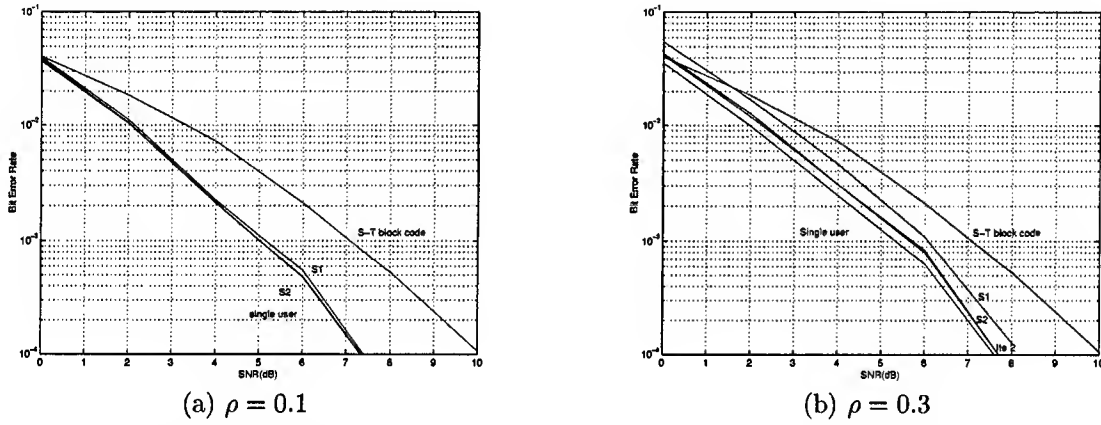


Figure 3: Performance of the multistage receiver for CDMA with Turbo coding and space-time block coding with  $K=4$  users, 2 transmit and 2 receive antennas.

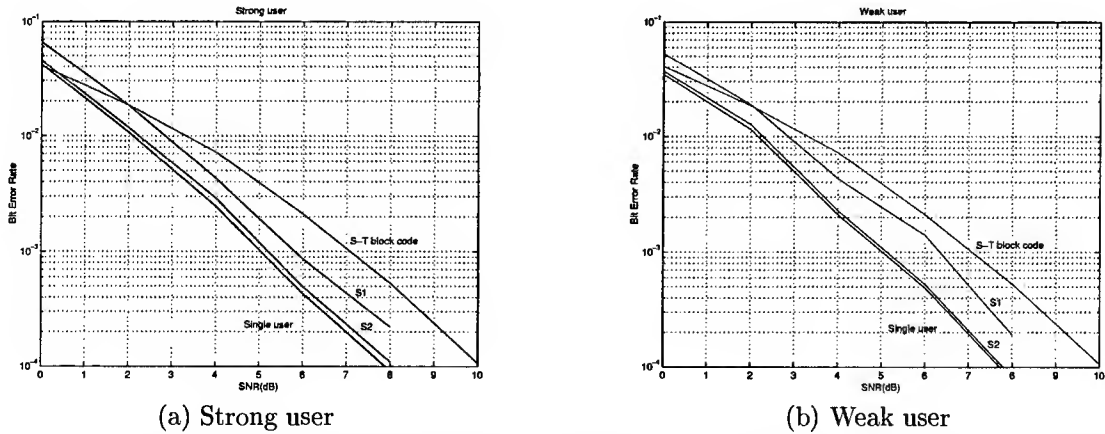


Figure 4: Performance of the multistage receiver for CDMA with Turbo coding and space-time block coding under a near-far situation with  $K=4$ ,  $\rho = 0.3$ , 2 transmit and 2 receive antennas. Two users are 20dB stronger than the other two users.

# ADAPTIVE MAP MULTI-USER DETECTION FOR FADING CDMA CHANNELS

Christophe Andrieu<sup>†</sup> - Arnaud Doucet<sup>†</sup> - Azzedine Touzni<sup>‡</sup>

<sup>†</sup>Signal Processing Group, Engineering Dept. Cambridge University  
Trumpington Street, CB2 1PZ Cambridge, UK.

<sup>‡</sup>NxtWave Communications, Langhorne, PA 19047, USA.

ca226@eng.cam.ac.uk - ad2@eng.cam.ac.uk - atouzni@nxtwavecomm.com

## ABSTRACT

This paper presents an adaptive multi-user maximum *a posteriori* (MAP) decoder for synchronous code division multiple access (CDMA) signals on fading channels. The key idea is to interpret this problem as an optimal filtering problem. An efficient particle filtering method is then developed to solve this complex estimation problem. Simulation results demonstrate the efficiency of our method.

## 1 Introduction

Code division multiple access (CDMA) systems have received much attention in recent years [13]. For the case of a known channel with additive Gaussian noise, the maximum likelihood (ML) optimal receiver was presented by Verdu [16]. Lower-complexity linear receivers have also been presented in this case. In the presence of unknown fading channels, the estimation problem to be solved is much more complex. MMSE linear receivers have also been presented in this context. However it turns out that the rate of adaptation for these linear techniques is not sufficient to track fast-fading channels and more sophisticated approaches are required. Recently, more efficient methods have been proposed; see for example [5], [6] where coupled estimators combining a Viterbi algorithm and an MMSE predictor are presented.

In this paper we follow a Bayesian probabilistic approach. A state-space model is included to model explicitly the non-stationarity of the fading channel. This allows us to formulate the problem of estimating *a posteriori* symbol probabilities as a complex optimal filtering problem. Under assumptions detailed later on, it is well known that exact computation of these probabilities involves a prohibitive computational cost exponential in the (growing) number of observations. Thus one needs to perform some approximations.

We present here a simulation-based method for solving this problem. This so-called *particle filtering* method can be viewed as a randomized adaptive grid approximation of the posterior distribution. As will be shown later, the particles

(values of the grid) evolve randomly in time according to a simulation-based rule. The weights of the particles are updated according to Bayes' rule. The most striking advantage of these MC particle filters is that the rate of convergence of the error towards zero is independent of the state dimension. That is, the randomization implicit in the particle filter gets around the curse of dimensionality. Taking advantage of the increase of computational power and the availability of parallel computers, several authors have recently proposed such particle methods following the seminal paper of Gordon *et al.* [11], see [7], [8] for a summary of the state-of-the-art and [2], [14], [15] for other applications in digital communications. It has been shown that these methods outperform the standard suboptimal methods.

We propose in this paper an improved particle method where the filtering distribution of interest is approximated by a Gaussian mixture of a large number, say  $N$ , of components which evolve stochastically over time and are driven by the observations. Though it is rather computationally intensive, it can be easily implemented on parallel processors.

The rest of the paper is organized as follows. In Section 2, we state the model and the estimation objectives. In Section 3, we describe particle filtering methods. Finally we demonstrate the efficiency of our algorithm in Section 4.

## 2 System Model and Estimation Objectives

### 2.1 System model

We follow here the presentation in [5], [6]. Consider a synchronous CDMA system with a single-antenna at the centralized receiver. The system has  $M$  users, each transmitting using a known direct sequence (DS) spreading code with processing gain  $G$  (i.e.  $G$  chips per symbol). For user  $m$ , the spreading code is represented by the  $G \times 1$  vector  $\mathbf{s}_m = [s_{m,0}, \dots, s_{m,G-1}]^T$ . At time  $t$ , user  $m$  transmits a symbol  $x_{m,t}$  of period  $T = GT_c$ , where  $T_c$  is the chip interval. Each chip  $s_{m,c}x_{m,t}$  is affected by the flat-fading channel  $f_{m,k}$ , represented at the chip rate where  $k = Gt + c$ . Note that  $t$  is used as an index at the symbol rate, and  $k$  is used as an index at the chip rate.

C. Andrieu is sponsored by AT&T Laboratories, Cambridge UK.

At the receiver, the incoming signal is sampled at the chip rate to obtain  $z_k$ . Assuming a synchronous system, the received samples are given by

$$z_k = \sum_{m=1}^M x_{m, \lfloor k/G \rfloor} s_{m, \text{mod}(k, G)} f_{m, k} + w_k$$

for  $k = 0, \dots, GT - 1$ . In vector-matrix notation

$$\mathbf{z}_t = \sum_{m=1}^M x_{m, t} \mathbf{S}_m \mathbf{f}_{m, t} + \mathbf{w}_t, \quad (1)$$

for  $t = 0, \dots, T - 1$ , where  $\mathbf{S}_m = \text{diag}(\mathbf{s}_m)$ ,  $\mathbf{z}_t \triangleq [z_{Gt}, \dots, z_{G(t+1)-1}]^T$ ,  $\mathbf{w}_t \triangleq [w_{Gt}, \dots, w_{G(t+1)-1}]^T$  is a vector of zero mean i.i.d. complex Gaussian noise samples with variance  $\sigma_w^2 = \frac{1}{2} \mathbb{E}[w_k w_k^*] = N_0 / (2T_c)$ . We assume that the fading channels  $\mathbf{f}_{m, t}$  satisfy the following state-space models

$$\mathbf{f}_{m, t} = \mathbf{A} \mathbf{f}_{m, t-1} + \mathbf{B} \mathbf{v}_{m, t} \quad (2)$$

where  $\mathbf{f}_{m, 0}$  is assumed distributed according to a Gaussian distribution and the disturbance noise  $\mathbf{v}_{m, t}$  is assumed zero mean i.i.d. Gaussian. We denote  $\mathbf{f}_t \triangleq [\mathbf{f}_{1, t}, \dots, \mathbf{f}_{M, t}]$ . The initial states  $\mathbf{f}_{m, 0}$ , the sequences  $\mathbf{v}_{m, t}$  and the observation noise  $\mathbf{w}_t$  are all assumed mutually independent at any time  $t$ . Finally, we assume that the symbols  $\mathbf{x}_t$  are modeled as a first-order (finite state-space) Markov chain. The finite state-space of the symbols is denoted by  $X$ .

## 2.2 Estimation objectives

Given the observations  $\mathbf{z}_{0:t} \triangleq (\mathbf{z}_0, \dots, \mathbf{z}_t)$ , all Bayesian inference on  $\mathbf{x}_{0:t} \triangleq (\mathbf{x}_0, \dots, \mathbf{x}_t)$  and  $\mathbf{f}_{0:t} \triangleq (\mathbf{f}_0, \dots, \mathbf{f}_t)$  is based on the posterior distribution  $p(\mathbf{x}_{0:t}, \mathbf{f}_{0:t} | \mathbf{z}_{0:t})$ . Here the channel coefficients  $\mathbf{f}_t$  are regarded as nuisance parameters and integrated out.

Our aim is to compute *recursively in time  $t$*  the MMAP symbol estimate defined as

$$\mathbf{x}_t^{\text{MMAP}} = \arg \max_{\mathbf{x}_t} p(\mathbf{x}_t | \mathbf{z}_{0:t})$$

The joint distribution  $p(\mathbf{x}_{0:t} | \mathbf{z}_{0:t})$  satisfies the following recursion

$$p(\mathbf{x}_{0:t+1} | \mathbf{z}_{0:t+1}) = p(\mathbf{x}_{0:t} | \mathbf{z}_{0:t}) \times \frac{p(\mathbf{z}_{t+1} | \mathbf{z}_{0:t}, \mathbf{x}_{0:t+1}) p(\mathbf{x}_{t+1} | \mathbf{x}_t)}{p(\mathbf{z}_{t+1} | \mathbf{z}_{0:t}, \mathbf{x}_{0:t})}.$$

The likelihood term  $p(\mathbf{z}_{t+1} | \mathbf{z}_{0:t}, \mathbf{x}_{0:t+1})$  can be evaluated pointwise through the Kalman filter associated to the path  $\mathbf{x}_{0:t+1}$  as the system (1)-(2) is linear Gaussian conditional upon  $\mathbf{z}_{0:t}$ . It is easily seen that, given our assumptions, computing  $p(\mathbf{x}_{0:t} | \mathbf{z}_{0:t})$  or  $p(\mathbf{x}_t | \mathbf{z}_{0:t})$  requires a computational

cost exponential in the (growing) number  $t$  of observations. It is thus necessary to develop an approximation scheme.

Efficient batch algorithms have been developed to solve related estimation problems [9] but they are of limited interest in a digital communications framework. Several "classical" suboptimal algorithms have also been proposed to solve related problems in the literature, see for example [1] for a standard textbook on the subject. However, these approximation methods are notoriously unreliable and faults are difficult to diagnose on-line.

## 3 Particle Filtering

In this paper, we present an original particle filtering method to solve this optimal estimation problem.

### 3.1 Perfect Monte Carlo sampling

Assume it is possible to sample  $N$  i.i.d. samples, called particles,  $\{\mathbf{x}_{0:t}^{(i)} : i = 1, \dots, N\}$  according to the joint distribution  $p(\mathbf{x}_{0:t} | \mathbf{y}_{1:t})$ , then an empirical distribution approximation of  $p(\mathbf{x}_{0:t} | \mathbf{y}_{1:t})$  is given by

$$p_N(\mathbf{x}_{0:t} | \mathbf{z}_{0:t}) = \frac{1}{N} \sum_{i=1}^N \delta_{\mathbf{x}_{0:t}^{(i)}}(\mathbf{x}_{0:t}).$$

Consequently an approximation of its marginal  $p(\mathbf{x}_t | \mathbf{z}_{0:t})$  is given by

$$p_N(\mathbf{x}_t | \mathbf{z}_{0:t}) = \frac{1}{N} \sum_{i=1}^N \delta_{\mathbf{x}_t^{(i)}}(\mathbf{x}_t)$$

that is, for any  $i \in X$ ,

$$p_N(\mathbf{x}_t = i | \mathbf{z}_{0:t}) = \frac{1}{N} \sum_{i=1}^N \delta_{\mathbf{x}_t^{(i)}}(i) \quad (3)$$

and

$$\hat{\mathbf{x}}_t^{\text{MMAP}} = \arg \max_{\mathbf{x}_t \in X} p_N(\mathbf{x}_t | \mathbf{z}_{0:t})$$

The estimate (3) is unbiased and from the strong law of large numbers (SLLN),  $p_N(\mathbf{x}_t = i | \mathbf{z}_{0:t}) \rightarrow p(\mathbf{x}_t = i | \mathbf{z}_{0:t})$  almost surely as  $N \rightarrow +\infty$ . A central limit theorem (CLT) holds too. The main advantage of Monte Carlo methods over other numerical integration methods is that the rate of convergence of  $p_N(\mathbf{x}_t = i | \mathbf{z}_{0:t})$  towards  $p(\mathbf{x}_t = i | \mathbf{z}_{0:t})$  is independent of the dimension  $t$ . Unfortunately, it is not possible to sample directly from the distribution  $p(\mathbf{x}_{0:t} | \mathbf{z}_{0:t})$  at any  $t$ , and alternative strategies need to be investigated.

### 3.2 Sequential Bayesian Importance Sampling

An alternative solution to estimate  $p(\mathbf{x}_{0:t} | \mathbf{z}_{0:t})$  consists of using the importance sampling method. Suppose that  $N$  i.i.d. samples  $\{\mathbf{x}_{0:t}^{(i)} : i = 1, \dots, N\}$  can be easily simulated according to an arbitrary importance distribution  $\pi(\mathbf{x}_{0:t} | \mathbf{z}_{0:t})$ ,

such that  $p(\mathbf{x}_{0:t}|\mathbf{z}_{0:t}) > 0$  implies  $\pi(\mathbf{x}_{0:t}|\mathbf{z}_{0:t}) > 0$ . Using this distribution a Monte Carlo estimate of  $p(\mathbf{x}_t|\mathbf{z}_{0:t})$  may be obtained as

$$\hat{p}_N(\mathbf{x}_t = i|\mathbf{z}_{0:t}) = \sum_{i=1}^N \tilde{w}_{0:t}^{(i)} \delta_{\mathbf{x}_t^{(i)}}(i), \quad (4)$$

where  $w_{0:t}^{(i)} \propto w(\mathbf{x}_{0:t}^{(i)})$  ( $\sum_{i=1}^N w_{0:t}^{(i)} = 1$ ), is the normalised version of the importance weight  $w(\mathbf{x}_{0:t}^{(i)})$  defined as

$$w(\mathbf{x}_{0:t}^{(i)}) \propto \frac{p(\mathbf{x}_{0:t}^{(i)}|\mathbf{z}_{0:t})}{\pi(\mathbf{x}_{0:t}^{(i)}|\mathbf{z}_{0:t})}.$$

According to the SLLN,  $\hat{p}_N(\mathbf{x}_t = i|\mathbf{z}_{0:t})$  converges almost surely towards  $p(\mathbf{x}_t = i|\mathbf{z}_{0:t})$  as  $N \rightarrow +\infty$ , and under additional assumptions a CLT also holds.

The method described up to now is a batch method. In order to obtain the estimate of  $p(\mathbf{x}_{0:t}|\mathbf{z}_{0:t})$  sequentially, one should be able to propagate this estimate in time without modifying subsequently the past simulated trajectories  $\{\mathbf{x}_{0:t}^{(i)} : i = 1, \dots, N\}$ . This means that  $\pi(\mathbf{x}_{0:t}|\mathbf{z}_{0:t})$  should admit  $\pi(\mathbf{x}_{0:t-1}|\mathbf{z}_{0:t-1})$  as marginal distribution:

$$\pi(\mathbf{x}_{0:t}|\mathbf{z}_{0:t}) = \pi(\mathbf{x}_{0:t-1}|\mathbf{z}_{0:t-1})\pi(\mathbf{x}_t|\mathbf{z}_{0:t}, \mathbf{x}_{0:t-1}),$$

and the importance weights  $w(\mathbf{x}_{0:t})$  can then be evaluated recursively, *i.e.*

$$w(\mathbf{x}_{0:t}) = w(\mathbf{x}_{0:t-1}) \times w_t, \quad (5)$$

where

$$w_t = \frac{p(\mathbf{x}_t|\mathbf{z}_{0:t}, \mathbf{x}_{0:t-1})}{\pi(\mathbf{x}_t|\mathbf{z}_{0:t}, \mathbf{x}_{0:t-1})}.$$

There are an unlimited number of choices for the importance distribution  $\pi(\mathbf{x}_{0:t}|\mathbf{z}_{0:t})$ , the only restriction being that its support includes that of  $p(\mathbf{x}_{0:t}|\mathbf{z}_{0:t})$ . A sensible selection criterion is to choose a proposal that minimises the variance of the importance weights given  $\mathbf{x}_{0:t-1}$  and  $\mathbf{z}_{0:t}$ . The importance distribution that satisfies this condition is  $\pi(\mathbf{x}_t|\mathbf{z}_{0:t}, \mathbf{x}_{0:t-1}) = p(\mathbf{x}_t|\mathbf{z}_{0:t}, \mathbf{x}_{0:t-1})$ , and this “optimal” importance distribution is employed throughout the paper (see [7] for details).

### 3.3 Selection step

For importance distributions of the form specified by (5) the variance of the importance weights can only increase (stochastically) over time [7]. It is thus impossible to avoid a degeneracy phenomenon. Practically, after a few iterations of the algorithm, all but one of the normalised importance weights are very close to zero, and a large computational effort is devoted to updating trajectories whose contribution to the final estimate is almost zero. To avoid this, it is of crucial importance to include a selection step in the algorithm, the purpose of which is to discard particles

with low normalised importance weights and multiply those with high normalised importance weights. The weights of the “surviving” particles are reset to  $1/N$ . A selection procedure associates with each particle, say  $\tilde{\mathbf{x}}_{0:t}^{(i)}$ ,  $i = 1, \dots, N$ , a number of children  $N_i \in \mathbb{N}$ , such that  $\sum_{i=1}^N N_i = N$ , to obtain  $N$  new particles  $\{\mathbf{x}_{0:t}^{(i)} : i = 1, \dots, N\}$ . If  $N_i = 0$  then  $\tilde{\mathbf{x}}_{0:t}^{(i)}$  is discarded, otherwise it has  $N_i$  children at time  $t + 1$ . In this paper, the selection step is done according to a stratified sampling scheme [12], though other methods such as sampling importance resampling (SIR) [11] may be employed. The stratified sampling scheme proceeds as follows: generate  $N$  points equally spaced in the interval  $[0, 1]$ , and associate for each particle  $i$ , a number of children  $N_i$  equal to the number of points lying between the partial sums of weights  $q_{i-1}$  and  $q_i$ , where  $q_i = \sum_{j=1}^i \tilde{w}_t^{(j)}$  ( $\tilde{w}_t^{(j)} = [\sum_{i=1}^N \tilde{w}_t^{(i)}]^{-1} w_t^{(j)}$ ). This algorithm is such that  $\mathbb{E}[N_i] = N\tilde{w}_t^{(i)}$  and  $\text{var}[N_i] = \{N\tilde{w}_t^{(i)}\} \left(1 - \{N\tilde{w}_t^{(i)}\}\right)$  where, for any  $\alpha$ ,  $[\alpha]$  is the integer part of  $\alpha$  and  $\{\alpha\} \triangleq \alpha - [\alpha]$ .

#### 3.3.1 Algorithm

Given at time  $t - 1$ ,  $N \in \mathbb{N}^*$  random samples  $\mathbf{x}_{0:t-1}^{(i)}$  ( $i = 1, \dots, N$ ) distributed according to  $p(\mathbf{x}_{0:t-1}|\mathbf{z}_{0:t-1})$ , the MC filter proceeds as follows at time  $t$ .

#### Particle Filtering Algorithm

##### Sequential Importance Sampling step

- For  $i = 1, \dots, N$ , sample  $\tilde{\mathbf{x}}_t^{(i)} \sim \pi(\mathbf{x}_t|\mathbf{z}_{0:t}, \mathbf{x}_{0:t-1}^{(i)})$  and  $\tilde{\mathbf{x}}_{0:t}^{(i)} \triangleq (\mathbf{x}_{0:t-1}^{(i)}, \tilde{\mathbf{x}}_t^{(i)})$ .
- For  $i = 1, \dots, N$ , evaluate the importance weights up to a normalising constant:

$$w_t^{(i)} \propto \frac{p(\mathbf{z}_t|\mathbf{z}_{0:t-1}, \tilde{\mathbf{x}}_{0:t}^{(i)}) p(\tilde{\mathbf{x}}_t^{(i)}|\tilde{\mathbf{x}}_{0:t-1}^{(i)})}{\pi(\tilde{\mathbf{x}}_t^{(i)}|\mathbf{z}_{0:t}, \tilde{\mathbf{x}}_{0:t-1}^{(i)})}$$

and normalise them  $\tilde{w}_t^{(i)} \propto w_t^{(i)}$ ,  $\sum_{j=1}^N \tilde{w}_t^{(j)} = 1$ .

##### Selection step

- Multiply/Discard particles  $(\tilde{\mathbf{x}}_{0:t}^{(i)}; i = 1, \dots, N)$  with respect to high/low normalised importance weights  $\tilde{w}_t^{(i)}$  to obtain  $N$  particles  $(\mathbf{x}_{0:t}^{(i)}; i = 1, \dots, N)$ .

Clearly, the computational complexity of the proposed algorithm at each iteration is  $O(N)$ . Moreover, since the optimal and prior importance distributions  $\pi(\mathbf{x}_t|\mathbf{z}_{0:t}, \mathbf{x}_{0:t-1})$  and the associated importance weights depend on  $\mathbf{x}_{0:t-1}$  via

a set of low-dimensional sufficient statistics, only these values need to be kept in memory and, thus, the storage requirements for the proposed algorithm are also  $O(N)$  and do not increase over time.

### 3.3.2 Convergence Results

The following proposition is a straightforward consequence of Theorem 1 in [4], which itself is an extension of results in [3].

**Proposition 1** *For all  $t \geq 0$ , there exists  $c_t$  independent of  $N$  such that*

$$\mathbb{E} \left[ (p_N(\mathbf{x}_t = i | \mathbf{z}_{0:t}) - p(\mathbf{x}_t = i | \mathbf{z}_{0:t}))^2 \right] \leq \frac{c_t}{N}$$

The expectation operator is with respect to the randomness introduced in the particle filtering method. Though the particles are interacting, one observes that one keeps the “standard” rate of convergence of Monte Carlo methods.

### 4 Simulation Results

We demonstrate the performance of our multi-user MAP decoder for transmission of binary-shift-keyed (BPSK) symbols over fast fading CDMA channels. The simulation parameters were as follows:  $M = 3$ ,  $G = 10$  and a flat fading channel with fading rate  $0.05/T$ . We compared our results with [6] and the case where the channel is assumed known exactly. The results in terms of Bit Error Rate (BER) are presented in Fig. 1. We notice that when the SNR is large, our stochastic algorithm outperforms substantially that of [6]. Their deterministic algorithm can indeed get trapped in severe local maxima as the posterior distribution is peakier.

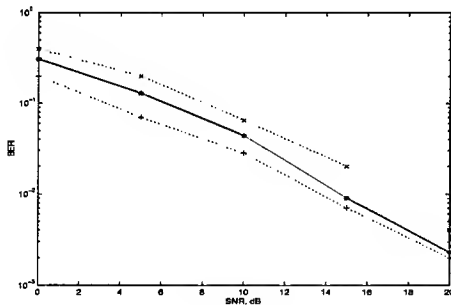


Figure 1: Dotted line + (channel known), solid line (particle filtering), dotted line x ([6])

### 5 REFERENCES

- [1] B.D.O. Anderson and J.B. Moore, Optimal Filtering, Prentice-Hall, Englewood Cliffs, 1979.
- [2] C. Andrieu, A. Doucet and E. Punskeya, “Sequential Monte Carlo methods for optimal filtering”, in [7].

- [3] D. Crisan, P. Del Moral and T. Lyons, “Discrete filtering using branching and interacting particle systems”, Markov Processes and Related Fields, vol. 5, no. 3, pp. 293-318, 1999.
- [4] D. Crisan and A. Doucet, “Convergence of generalized particle filters”, technical report, Cambridge University, TR-F-INFENG TR 381, 2000.
- [5] L.M. Davis and I.B. Collings, “Joint MAP detection and channel estimation for CDMA over frequency-selective fading channels”, in Proc. ISPACS-98, pp. 432-436, 1998.
- [6] L.M. Davis and I.B. Collings, “Multi-user MAP decoding for flat-fading CDMA channels”, in Proc. Conf. DSPCS-99, pp. 79-86, 1999.
- [7] A. Doucet, J.F.G. de Freitas and N.J. Gordon (eds.), Sequential Monte Carlo Methods in Practice, Springer-Verlag: New-York, 2000.
- [8] A. Doucet, S.J. Godsill and C. Andrieu, “On sequential Monte Carlo sampling methods for Bayesian filtering”, Statistics and Computing, vol. 10, no. 3, pp. 197-208, 2000.
- [9] A. Doucet, A. Logothetis and V. Krishnamurthy, “Stochastic sampling algorithms for state estimation of jump Markov linear systems”, IEEE Trans. Automatic Control, vol. 45, no. 2, pp. 188-201, 2000.
- [10] A. Doucet, N.J. Gordon and V. Krishnamurthy, “Particle filters for state estimation of jump Markov linear systems”, technical report, Cambridge University, TR-F-INFENG TR 359, 1999.
- [11] N.J. Gordon, D.J. Salmond and A.F.M. Smith, “Novel approach to nonlinear/non-Gaussian Bayesian state estimation”, IEE Proceedings-F, vol. 140, no. 2, pp. 107-113, 1993.
- [12] G. Kitagawa, “Monte Carlo Filter and Smoother for Non-Gaussian Nonlinear State Space Models”, J. Comp. Graph. Stat., vol. 5, no. 1, pp. 1-25, 1996.
- [13] U. Madhow, “Blind adaptive interference suppression for direct-sequence CDMA”, Proceedings of the IEEE, pp. 2049-2069, 1998.
- [14] E. Punskeya, C. Andrieu, A. Doucet and W.J. Fitzgerald, “Particle filters for demodulation of M-ary modulated signals in noisy fading communication channels”, in Proceedings Conf. ICASSP 2000.
- [15] E. Punskeya, C. Andrieu, A. Doucet and W.J. Fitzgerald, “Particle filtering for demodulation in fading channels”, technical report Cambridge University CUED-F-INFENG TR 381, 2000.
- [16] S. Verdu, “Minimum probability of error for asynchronous Gaussian multiple access channels”, IEEE Trans. Information Theory, vol. 32, no. 1, pp. 85-96, 1986.

# ANALYSIS OF A SUBSPACE CHANNEL ESTIMATION TECHNIQUE FOR MULTICARRIER CDMA SYSTEMS

*Carlos J. Escudero, Daniel I. Iglesia, Mónica F. Bugallo, Luis Castedo*

Departamento de Electrónica y Sistemas. Universidad de La Coruña  
Campus de Elviña s/n, 15.071 La Coruña, SPAIN  
Tel: ++ 34-981-167150, e-mail: escudero@des.fi.udc.es

## ABSTRACT

In this paper we investigate a blind channel estimation method for Multi-Carrier CDMA systems that uses a subspace decomposition technique. This technique exploits the orthogonality property between the noise subspace and the received user codes to obtain a channel identification algorithm. In order to analyze the performance of this algorithm, we derived a theoretical expression of the estimation MSE using a perturbation approach. This expression is compared with the numerical results of some computer simulations to illustrate the validity of the analysis.

## 1. INTRODUCTION

Multi-Carrier (MC) transmission methods for Code Division Multiple Access (CDMA) communication systems have been recently proposed as an efficient technique to combat multipath propagation and have gained an increased interest during the last years [1, 2]. In these techniques each user is assigned to a unique identification code sequence and the transmitted signal is split in different subcarriers. It is assumed that the subcarrier bandwidth is smaller than the channel coherence bandwidth and, therefore, presents only flat fading. As a consequence, MC-CDMA systems do not suffer from Inter-Symbol Interference (ISI). However, the effects of dispersive channels appear as random distortions in the amplitude and phase of each subcarrier. This causes a loss of orthogonality between user codes and introduces Multiple Access Interference (MAI).

In order to implement a multiuser detector and to reduce MAI it is necessary to characterize, implicitly or explicitly, the channel parameters. In this paper we introduce a new blind channel estimation technique that is based on a subspace decomposition [3] and derive a particular algorithm to identify the channel parameters. We also obtain, using perturbation techniques, an

approximate expression of the estimation Mean Square Error (MSE) achieved with the proposed algorithm.

The paper is organized as follows. Section 2 presents the signal model of a synchronous MC-CDMA system. Section 3 describes the subspace decomposition technique and the resultant algorithm. In section 4 we perform the theoretical analysis of the estimation MSE. Section 5 shows the results of several computer simulations that illustrate the validity of the approximations in the previous section and, finally, Section 6 is devoted to the conclusions.

## 2. SIGNAL MODEL

Let us consider a discrete-time baseband equivalent model of a synchronous MC-CDMA system with  $N$  users using  $L$ -chip signature codes. The  $k$ -th chip corresponding to the  $n$ -th symbol transmitted by the  $i$ -th user is given by

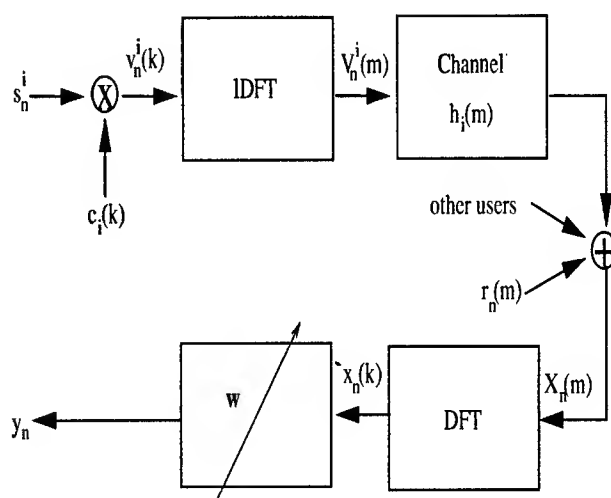


Figure 1: Block diagram of the discrete-time baseband model of a MC-CDMA system.

This work has been supported by FEDER (grant 1FD97-0082).

$$v_n^i(k) = s_n^i c_i(k) \quad k = 0, \dots, L-1 \quad n = 0, 1, 2, \dots \quad (1)$$

where  $c_i(k)$  is the  $k$ -th chip of the  $i$ -th user code. In a MC-CDMA system the modulator computes the  $L$ -IDFT (Inverse Discrete Fourier Transform) of (1) to obtain the following multicarrier signal

$$V_n^i(m) = IDFT[v_n^i(k)] = \frac{1}{L} \sum_{k=0}^{L-1} v_n^i(k) e^{j \frac{2\pi}{L} km} \quad (2)$$

This signal is transmitted through a dispersive channel with an impulse response  $h_i(m)$ ;  $m = 0, \dots, M-1$ . At the receiver the observed signal is a superposition of the signals corresponding to  $N$  users plus an additive white Gaussian noise (AWGN). Therefore, the received signal for the  $n$ -th symbol is the following

$$X_n(m) = \sum_{i=1}^N V_n^i(m) * h_i(m) + r_n(m) \quad (3)$$

where  $*$  denotes discrete convolution and  $r_n(m)$  represents a white noise sequence.

To recover the transmitted symbols, the receiver applies a  $L$ -DFT (Discrete Fourier Transform) to the received signal (3). Assuming perfect synchronization and a sufficiently large guard time between symbols, the resultant signal is

$$\begin{aligned} x_n(k) &= DFT[X_n(m)] = \sum_{i=1}^N v_n^i(k) H_i(k) + ,_n(k) \quad (4) \\ &= \sum_{i=1}^N s_n^i c_i(k) H_i(k) + ,_n(k) \quad k = 0, \dots, L-1 \end{aligned}$$

where  $H_i(k)$  and  $,_n(k)$  are the DFT's of  $h_i(m)$  and  $r_n(m)$ , respectively. Rewriting (4) in vector notation we obtain

$$\begin{aligned} \mathbf{x}_n &= [x_n(0), \dots, x_n(L-1)]^T = \sum_{i=1}^N s_n^i \mathbf{C}_i \mathbf{H}_i + \mathbf{\Gamma}_n \\ &= \sum_{i=1}^N s_n^i \mathbf{C}_i \mathbf{F} \mathbf{h}_i + \mathbf{\Gamma}_n = \sum_{i=1}^N s_n^i \tilde{\mathbf{c}}_i + \mathbf{\Gamma}_n \quad (5) \end{aligned}$$

where  $T$  denotes transposition,  $\mathbf{C}_i$  is a diagonal matrix whose elements are the  $L$  chips of the code corresponding to the  $i$ -th user,  $\mathbf{H}_i = [H_i(0), \dots, H_i(L-1)]^T$  and  $\mathbf{\Gamma}_n = [,_n(0), \dots, ,_n(L-1)]^T$ . To obtain (5) we have used the relationship  $\mathbf{H}_i = \mathbf{F} \mathbf{h}_i$  where  $\mathbf{F}$  is a  $L \times M$  DFT matrix and  $\mathbf{h}_i = [h_i(0), \dots, h_i(M-1)]^T$ . Note that (5) is a CDMA signal where the code associated to the  $i$ -th user is  $\tilde{\mathbf{c}}_i = \mathbf{C}_i \mathbf{F} \mathbf{h}_i$ .

### 3. SUBSPACE DECOMPOSITION

Assuming statistical independence between users and noise, the autocorrelation matrix of the observations vector (5) can be decomposed as

$$\begin{aligned} \mathbf{R} &= E[\mathbf{x}_n \mathbf{x}_n^H] = \sum_{i=1}^N \tilde{\mathbf{c}}_i E[s_n^i s_n^{i*}] \tilde{\mathbf{c}}_i^H + E[\mathbf{\Gamma}_n \mathbf{\Gamma}_n^H] \\ &= \sum_{i=1}^N \sigma_i^2 \tilde{\mathbf{c}}_i \tilde{\mathbf{c}}_i^H + \sigma_r^2 \mathbf{I} \quad (6) \end{aligned}$$

where  $E[\cdot]$  is the expectation operator,  $*$  represents conjugate,  $^H$  denotes conjugate transpose,  $\mathbf{I}$  is the identity matrix and  $\sigma_i^2$  and  $\sigma_r^2$  are the  $i$ -th user signal and noise power, respectively.

Let us consider the eigendecomposition of (6). There are  $L$  eigenvalues that we sort as  $\lambda_0 \geq \lambda_1 \geq \dots \geq \lambda_{L-1}$ . It is well-known that the eigenvectors associated to the  $N$  most significant eigenvalues ( $\mathbf{u}_l$ ,  $l = 0, \dots, N-1$ ) span the signal subspace where the perturbed user codes,  $\tilde{\mathbf{c}}_i$ , lie. The remaining  $L-N$  eigenvectors ( $\mathbf{u}_l$ ,  $l = N, \dots, L-1$ ) span the noise (orthogonal) subspace and their associated eigenvalues are equal to the noise power, i.e.,  $\lambda_N = \dots = \lambda_{L-1} = \sigma_r^2$  [3].

As we have seen, the perturbed user codes lie in the signal subspace and are orthogonal to the noise subspace. This property can be used to state the following system of equations for the  $i$ -th user

$$\tilde{\mathbf{c}}_i^H \mathbf{u}_l = 0 \quad l = N, \dots, L-1 \quad (7)$$

Recall that this system of equations has  $M$  unknowns and  $L-N$  equations. It will be solvable if and only if the number of equations is greater or equal than the number of unknowns,  $M \leq L-N$ . This means that the number of simultaneous users,  $N$ , is limited by the number of carriers,  $L$ , and the channel length,  $M$ . Nevertheless, it is interesting to note that the system capacity can be increased without increasing the number of carriers using codes with a length larger than the spreading gain [4].

In order to solve the equations system (7), we can consider the following equivalent system

$$\|\tilde{\mathbf{c}}_i^H \mathbf{u}_l\|^2 = \tilde{\mathbf{c}}_i^H \mathbf{u}_l \mathbf{u}_l^H \tilde{\mathbf{c}}_i = \mathbf{h}_i^H \mathbf{F}^H \mathbf{C}_i^H \mathbf{u}_l \mathbf{u}_l^H \mathbf{C}_i \mathbf{F} \mathbf{h}_i = 0 \quad (8)$$

for  $l = N, \dots, L-1$ . The solution to these equations can be found by solving the following minimization problem

$$\hat{\mathbf{h}}_i = \arg \min_{\|\mathbf{h}_i\|^2=1} \left[ \sum_{l=N}^{L-1} \mathbf{h}_i^H \mathbf{F}^H \mathbf{C}_i^H \mathbf{u}_l \mathbf{u}_l^H \mathbf{C}_i \mathbf{F} \mathbf{h}_i \right]$$

$$\begin{aligned}
&= \arg \min_{\|\mathbf{h}_i\|^2=1} \mathbf{h}_i^H \left[ \sum_{l=N}^{L-1} \mathbf{F}^H \mathbf{C}_i^H \mathbf{u}_l \mathbf{u}_l^H \mathbf{C}_i \mathbf{F} \right] \mathbf{h}_i \\
&= \arg \min_{\|\mathbf{h}_i\|^2=1} \mathbf{h}_i^H [\mathbf{F}^H \mathbf{C}_i^H \mathbf{U} \mathbf{U}^H \mathbf{C}_i \mathbf{F}] \mathbf{h}_i \\
&= \arg \min_{\|\mathbf{h}_i\|^2=1} \mathbf{h}_i^H \mathbf{Q}_i \mathbf{h}_i
\end{aligned} \tag{9}$$

where the solution  $\hat{\mathbf{h}}_i$  is an estimation of the channel impulse response vector,  $\mathbf{U}$  is a  $L \times (L - N)$  matrix whose columns are the eigenvectors associated to the noise subspace (i.e.,  $\mathbf{u}_l$ ,  $l = N, \dots, L - 1$ ) and  $\mathbf{Q}_i = \mathbf{F}^H \mathbf{C}_i^H \mathbf{U} \mathbf{U}^H \mathbf{C}_i \mathbf{F}$ . The solution can be obtained by the least squares method and it corresponds to the eigenvector of  $\mathbf{Q}_i$  associated to its minimum eigenvalue [5].

In practice, we do not know *a priori* the autocorrelation matrix (6). However, it can be estimated from the sampled matrix as

$$\hat{\mathbf{R}} = \frac{1}{N_s} \sum_{n=1}^{N_s} \mathbf{x}_n \mathbf{x}_n^H \tag{10}$$

where  $N_s$  is the number of received symbols used to obtain the estimation. Note that  $\hat{\mathbf{R}} \rightarrow \mathbf{R}$  as  $N_s$  tends to infinity and also its eigenvalues  $\hat{\lambda}_l \rightarrow \lambda_l$  and eigenvectors  $\hat{\mathbf{u}}_l \rightarrow \mathbf{u}_l$ .

Finally, when using second order statistics, the channel impulse response can be obtained up to a complex constant. This constant has to be compensated in order to analyze the algorithm performance. Towards this aim, we normalize the estimation of the impulse response vector as  $\hat{\mathbf{h}}_{i,normalized} = \frac{h_i(0)}{\hat{h}_i(0)} \hat{\mathbf{h}}_i$  where  $h_i(0)$  and  $\hat{h}_i(0)$  are the first elements of the true and estimated channel impulse response vectors, respectively.

#### 4. MEAN SQUARE ERROR ANALYSIS

In this section, we derive an analytical expression of the estimation MSE. For simplicity reasons, let us denote  $\mathbf{h}_i = \mathbf{h}$ ,  $\mathbf{Q}_i = \mathbf{Q}$  and  $\mathbf{C}_i = \mathbf{C}$ . Our analysis is based on a perturbation technique [7] that allows us to express the perturbation in  $\mathbf{h}$ ,  $\Delta \mathbf{h}$ , in terms of the perturbation in  $\mathbf{Q}$ ,  $\Delta \mathbf{Q}$ . Let us consider the following identities

$$\begin{aligned}
\mathbf{Q} \mathbf{h} &= \mathbf{0} \\
\hat{\mathbf{h}} &= \mathbf{h} + \Delta \mathbf{h} \\
\hat{\mathbf{Q}} &= \mathbf{Q} + \Delta \mathbf{Q}
\end{aligned} \tag{11}$$

For a sufficiently large number of samples ( $N_s \rightarrow \infty$ ),  $\hat{\mathbf{Q}} \rightarrow \mathbf{Q}$ ,  $\hat{\mathbf{h}} \rightarrow \mathbf{h}$  and  $\hat{\mathbf{Q}} \hat{\mathbf{h}}$  is approximately equal to the zero vector, i.e.

$$\hat{\mathbf{Q}} \hat{\mathbf{h}} = (\mathbf{Q} + \Delta \mathbf{Q})(\mathbf{h} + \Delta \mathbf{h}) \simeq \Delta \mathbf{Q} \mathbf{h} + \mathbf{Q} \Delta \mathbf{h} \approx \mathbf{0}$$

where we have neglected the second order term,  $\Delta \mathbf{Q} \Delta \mathbf{h} \simeq 0$ . Therefore,

$$\mathbf{Q} \Delta \mathbf{h} \simeq -\Delta \mathbf{Q} \mathbf{h} \tag{12}$$

and

$$\begin{aligned}
\Delta \mathbf{h} &\simeq -\mathbf{Q}^\dagger \Delta \mathbf{Q} \mathbf{h} \\
&= -\mathbf{Q}^\dagger (\hat{\mathbf{Q}} - \mathbf{Q}) \mathbf{h} \\
&= -\mathbf{Q}^\dagger \hat{\mathbf{Q}} \mathbf{h}
\end{aligned} \tag{13}$$

where  $\mathbf{Q}^\dagger$  denotes the left pseudo-inverse of  $\mathbf{Q}$ . The  $k$ -th component of  $\Delta \mathbf{h}$  is given by

$$\begin{aligned}
\Delta h(k) &\simeq -\mathbf{q}_k^H \hat{\mathbf{Q}} \mathbf{h} \\
&= -\mathbf{q}_k^H (\mathbf{F}^H \mathbf{C}^H \hat{\mathbf{U}} \hat{\mathbf{U}}^H \mathbf{C} \mathbf{F}) \mathbf{h} \\
&= -\sum_{l=N}^{L-1} \mathbf{q}_k^H \mathbf{F}^H \mathbf{C}^H \hat{\mathbf{u}}_l \hat{\mathbf{u}}_l^H \mathbf{C} \mathbf{F} \mathbf{h} \\
&= -\sum_{l=N}^{L-1} \hat{\mathbf{u}}_l^H \mathbf{C} \mathbf{F} \mathbf{h} \mathbf{q}_k^H \mathbf{F}^H \mathbf{C}^H \hat{\mathbf{u}}_l \\
&= -\text{Trace}\{\hat{\mathbf{U}} \hat{\mathbf{U}}^H \mathbf{C} \mathbf{F} \mathbf{h} \mathbf{q}_k^H \mathbf{F}^H \mathbf{C}^H\}
\end{aligned} \tag{14}$$

where  $\mathbf{q}_k$  is the  $k$ -th column of  $(\mathbf{Q}^\dagger)^H$ . Based on the results of [6] (page 1840, equation (4.11)), we obtain the following identity

$$\hat{\mathbf{U}} \hat{\mathbf{U}}^H \mathbf{C} \mathbf{F} \mathbf{h} \simeq -\mathbf{U} \mathbf{U}^H \Delta \mathbf{V} \mathbf{V}^H \mathbf{C} \mathbf{F} \mathbf{h} \tag{15}$$

where  $\mathbf{V}$  is a  $L \times N$  matrix whose columns are the eigenvectors associated to the signal subspace (i.e.,  $\mathbf{u}_l$ ,  $l = 0, \dots, N - 1$ ) and  $\Delta \mathbf{V} = \hat{\mathbf{V}} - \mathbf{V}$ . Moreover, from Appendix A of [6] (page 1844, equation (A.2))

$$\mathbf{U}^H \Delta \mathbf{V} \simeq \mathbf{U}^H \hat{\mathbf{R}} \mathbf{V} \mathbf{\Lambda}^{-1} \tag{16}$$

where  $\mathbf{\Lambda} = \text{diag}(\lambda_0 - \sigma_r^2, \dots, \lambda_{N-1} - \sigma_r^2)$  where  $\text{diag}(\mathbf{a})$  is a diagonal matrix whose elements are the elements of vector  $\mathbf{a}$ . To remove the effect of the unknown constant that we have in the estimation of the channel vector, we have to consider a normalization of the vector channel estimate. Similarly to [7], we select the following normalization

$$\Delta \mathbf{h}_{normalized} = (\mathbf{I} - \frac{\mathbf{h} \mathbf{1}^T}{h(0)}) \Delta \mathbf{h} \tag{17}$$

where  $\mathbf{I}$  is the identity matrix and  $\mathbf{1}^T = [1, 0, 0, \dots]$ . This normalization can be included in (13) and now  $\mathbf{q}_k$  will be the  $k$ -th column of the matrix  $((\mathbf{I} - \frac{\mathbf{h} \mathbf{1}^T}{h(0)}) \mathbf{Q}^\dagger)^H$ . Combining (15) and (16) in (14), we obtain the following expression

$$\Delta h(k) \simeq \text{Trace}\{\mathbf{U} \mathbf{U}^H \hat{\mathbf{R}} \mathbf{V} \mathbf{\Lambda}^{-1} \mathbf{V}^H \mathbf{C} \mathbf{F} \mathbf{h} \mathbf{q}_k^H \mathbf{F}^H \mathbf{C}^H\}$$



$$\begin{aligned}
&= \sum_{l=N}^{L-1} (\mathbf{u}_l^H \hat{\mathbf{R}} \mathbf{V} \mathbf{\Lambda}^{-1} \mathbf{V}^H \mathbf{C} \mathbf{F} \mathbf{h} \mathbf{q}_k^H \mathbf{F}^H \mathbf{C}^H \mathbf{u}_l) \\
&= \sum_{l=N}^{L-1} \mathbf{u}_l^H \hat{\mathbf{R}} \mathbf{g}_{lk}
\end{aligned} \tag{18}$$

where  $\mathbf{g}_{lk} = \mathbf{V} \mathbf{\Lambda}^{-1} \mathbf{V}^H \mathbf{C} \mathbf{F} \mathbf{h} \mathbf{q}_k^H \mathbf{F}^H \mathbf{C}^H \mathbf{u}_l$ .

Finally, to obtain the MSE of the channel estimation algorithm, we have to explore the fourth order statistics of binary and Gaussian random variables. In appendix A it is demonstrated that

$$\begin{aligned}
E[||\Delta \mathbf{h}||^2] &= \\
&= \frac{\sigma_r^2}{N_s} \sum_{k=0}^{M-1} (\text{Trace}\{\mathbf{U}^H \mathbf{U} \mathbf{G}_k^H \tilde{\mathbf{C}} \tilde{\mathbf{C}}^H \mathbf{G}_k\} \\
&\quad + \sigma_r^2 \text{Trace}\{\mathbf{U}^H \mathbf{U} \mathbf{G}_k^H \mathbf{G}_k\})
\end{aligned} \tag{19}$$

where  $\mathbf{G}_k = [\mathbf{g}_{Nk}, \dots, \mathbf{g}_{(L-1)k}]$  and  $\tilde{\mathbf{C}} = [\sigma_1 \tilde{\mathbf{c}}_1, \dots, \sigma_N \tilde{\mathbf{c}}_N]$ .

## 5. SIMULATIONS

In this section we compare the analytical expression (19) with the MSE obtained from computer simulations of the algorithm (9) to illustrate the validity of the approximation carried out in the previous section.

Figure 2 examines the accuracy of the MSE analysis. It is shown the time evolution for theoretical and simulated MSE (averaged value of 50 realizations). An environment with  $L = 12$  carriers, a channel length  $M = 4$  and 8 users received with a  $SNR = 12dB$  was considered. It can be seen that even for a small number of symbols, the theoretical expression fits to the simulated MSE.

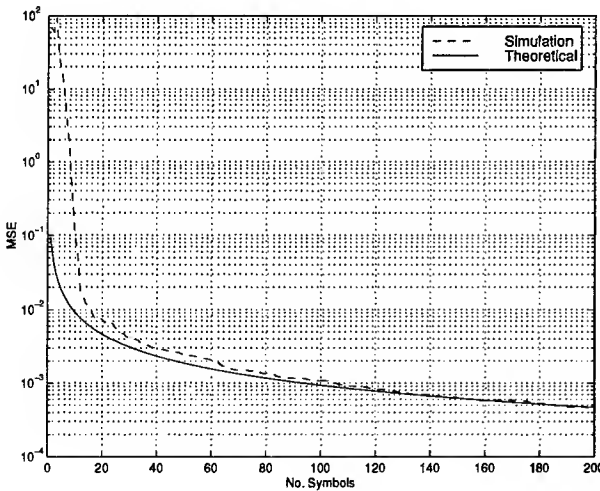


Figure 2: Time evolution of the simulated and theoretical MSE.

Figure 3 shows the simulated and theoretical MSE versus the Signal to Noise Ratio (SNR) of the received users. The environment is the same as before and the curves are obtained after  $N_s = 200$  symbols. We can see that both curves are very similar even for small values of SNR.

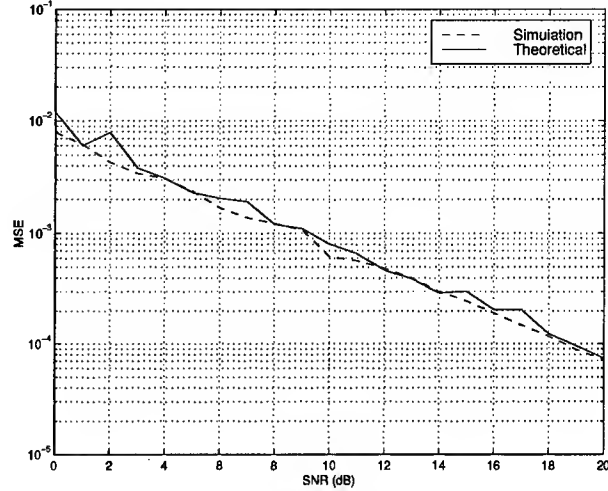


Figure 3: Simulated and theoretical MSE vs. received users SNR.

## 6. CONCLUSIONS

A new blind channel identification method for Multi-Carrier CDMA systems has been presented. The method exploits the orthogonality between the signal and noise subspaces of the incoming signal. It also has been investigated the performance of the method: using a perturbation technique, we derived an analytical approximate expression of the estimation MSE. Computer simulations have revealed the high accuracy of the analytical approximation carried out.

### A. APPENDIX

Taking into account that  $\tilde{\mathbf{c}}_i^H \mathbf{u}_l = \mathbf{u}_l^H \tilde{\mathbf{c}}_i = 0$ , it is straightforward to obtain from (18) that

$$\begin{aligned}
\Delta \mathbf{h}(k) &= \\
&= \frac{1}{N_s} \sum_{l=N}^{L-1} \sum_{n=0}^{N_s-1} \mathbf{u}_l^H \left( \sum_{i=1}^N \Gamma_n(s_n^i)^* \tilde{\mathbf{c}}_i^H + \Gamma_n \Gamma_n^H \right) \mathbf{g}_{lk}
\end{aligned} \tag{20}$$

where  $*$  represents conjugate. Therefore, the MSE is

$$E[||\Delta \mathbf{h}||^2] = \sum_{k=0}^{M-1} E[\Delta \mathbf{h}(k) \Delta \mathbf{h}^*(k)] = \tag{21}$$

$$\begin{aligned}
&= \sum_{k=0}^{M-1} \left( \frac{1}{N_s^2} \sum_{l=N}^{L-1} \sum_{p=N}^{L-1} \sum_{n=0}^{N_s-1} \sum_{m=0}^{N_s-1} \sum_{i=1}^N \sum_{j=1}^N \right. \\
&\quad \left. E[\mathbf{u}_l^H \Gamma_n \Gamma_m^H \mathbf{u}_p \mathbf{g}_{pk}^H \tilde{\mathbf{c}}_i s_m^i (s_n^j)^* \tilde{\mathbf{c}}_j^H \mathbf{g}_{lk}] \right. \\
&\quad \left. + \frac{1}{N_s^2} \sum_{l=N}^{L-1} \sum_{p=N}^{L-1} \sum_{n=0}^{N_s-1} \sum_{m=0}^{N_s-1} E[\mathbf{u}_l^H \Gamma_n \Gamma_n^H \mathbf{g}_{lk} \mathbf{g}_{pk}^H \Gamma_m \Gamma_m^H \mathbf{u}_p] \right)
\end{aligned}$$

where we have used the fact that the third order moments of a Gaussian random variable are zero.

Considering statistical independence between users and noise and the user symbols i.i.d., the first expectation (21) is

$$\begin{aligned}
&E[\mathbf{u}_l^H \Gamma_n \Gamma_m^H \mathbf{u}_p \mathbf{g}_{pk}^H \tilde{\mathbf{c}}_i s_m^i (s_n^j)^* \tilde{\mathbf{c}}_j^H \mathbf{g}_{lk}] = \\
&= \mathbf{u}_l^H E[\Gamma_n \Gamma_m^H] \mathbf{u}_p \mathbf{g}_{pk}^H \tilde{\mathbf{c}}_i E[s_m^i (s_n^j)^*] \tilde{\mathbf{c}}_j^H \mathbf{g}_{lk} \\
&= \sigma_i^2 \sigma_r^2 \mathbf{u}_l^H \mathbf{u}_p \mathbf{g}_{pk}^H \tilde{\mathbf{c}}_i \tilde{\mathbf{c}}_j^H \mathbf{g}_{lk} \delta(n-m) \delta(i-j) \quad (22)
\end{aligned}$$

where  $\delta(\cdot)$  is the Kronecker function.

The second expectation in (21) can be expressed as

$$\begin{aligned}
&E[\mathbf{u}_l^H \Gamma_n \Gamma_n^H \mathbf{g}_{lk} \mathbf{g}_{pk}^H \Gamma_m \Gamma_m^H \mathbf{u}_p] = \\
&= \mathbf{u}_l^H E[\Gamma_n \Gamma_n^H] \mathbf{g}_{lk} \mathbf{g}_{pk}^H E[\Gamma_m \Gamma_m^H] \mathbf{u}_p \\
&\quad + \mathbf{u}_l^H E[\Gamma_n \Gamma_m^H] \mathbf{u}_p \mathbf{g}_{pk}^H E[\Gamma_m \Gamma_n^H] \mathbf{g}_{lk} \\
&= \sigma_r^4 \mathbf{u}_l^H \mathbf{u}_p \mathbf{g}_{pk}^H \mathbf{g}_{lk} \delta(n-m) \quad (23)
\end{aligned}$$

where we have used the facts  $\mathbf{u}_l^H \mathbf{g}_{lk} = 0$  and  $E[\theta_1 \theta_2^* \theta_3 \theta_4^*] = E[\theta_1 \theta_2^*] E[\theta_3 \theta_4^*] + E[\theta_1 \theta_4^*] E[\theta_2 \theta_3^*]$  when  $\theta_i$   $i = 1, 2, 3, 4$  are four independent Gaussian variables [7].

Including (22) and (23) in (21), it is obtained

$$\begin{aligned}
&E[||\Delta \mathbf{h}||^2] = \\
&= \frac{1}{N_s^2} \sum_{k=0}^{M-1} \sum_{l=N}^{L-1} \sum_{p=N}^{L-1} \sum_{n=0}^{N_s-1} \sum_{i=1}^N \left( \sum_{m=0}^{N_s-1} \sigma_r^2 \sigma_i^2 \mathbf{u}_l^H \mathbf{u}_p \mathbf{g}_{pk}^H \tilde{\mathbf{c}}_i \tilde{\mathbf{c}}_i^H \mathbf{g}_{lk} \right. \\
&\quad \left. + \sigma_r^4 \mathbf{u}_l^H \mathbf{u}_p \mathbf{g}_{pk}^H \mathbf{g}_{lk} \right) \quad (24)
\end{aligned}$$

that is equivalent to (19).

## REFERENCES

- [1] K. Fazel, G. P. Fettweis, *Multi-Carrier Spread-Spectrum*, Kluwer Academic Publishers, 1997.
- [2] N. Yee, J. P. Linnartz, G. Fettweis, "Multi-Carrier CDMA in Indoor Wireless Radio Networks", *Proc. International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC93)*, Yokohama, pp. 109-113, 1993.
- [3] E. Moulines, P. Duhamel, J. F. Cardoso and S. Mayrargue, "Subspace Methods for the Blind

Identification of Multichannel FIR Filters", *IEEE Transactions on Signal Processing*, vol. 43, no. 2, pp. 516-525, February 1995.

- [4] D. I. Iglesias, C. J. Escudero, L. Castedo, "A Subspace Method for Blind Channel Identification in Multi-Carrier CDMA Systems", *Second International Workshop on Multi-Carrier Spread Spectrum & Related Topics (MCSS'99)*, Kluwer Academic Publishers, September 1999.
- [5] G. Strang, *Linear Algebra and its Applications*, Harcourt Brace Jovanovich, Third Edition, 1988.
- [6] P. Stoica and T. Söderström, "Statistical Analysis and Subspace Rotation Estimates of Sinusoidal Frequencies", *IEEE Transactions on Signal Processing*, vol. 39, no. 8, pp.1836-1847, August 1991.
- [7] W. Qiu, Y. Hua, "Performance Analysis of the Subspace Method for Blind Channel Identification", *Signal Processing*, no. 50, pp. 71-81, 1996.

# BLIND ADAPTIVE ASYNCHRONOUS CDMA MULTIUSER DETECTOR USING PREDICTION LEAST MEAN KURTOSIS ALGORITHM

*Kunjie Wang and Yeheskel Bar-Ness*

Center for Communications and Signal Processing Research  
Department of Electrical and Computer Engineering  
New Jersey Institute of Technology  
University Heights, Newark, NJ 07102, USA  
Tel: 1-973-596-3520 Fax: 1-973-596-8473  
Email: wangk@njit.edu Cc: barness@njit.edu

## ABSTRACT

In this paper, a new blind adaptive multiuser detector, which is termed prediction least mean kurtosis (PLMK) algorithm, is proposed for joint MAI and narrowband interference (NBI) suppression in asynchronous CDMA systems. This algorithm is based on a higher-order statistics rather than the second-order statistics used in the LMS algorithm. Unlike the regular least mean kurtosis (LMK), it takes into consideration samples earlier than those correspond to current bit. For comparison purposes, we also apply the regular LMK algorithm to the case of asynchronous CDMA systems. Simulation results show that the blind adaptive multiuser detector with PLMK algorithm provides significantly better performance than the one with regular LMK algorithm.

## 1. INTRODUCTION

Blind adaptive multiuser detector has received significant attention due to its implementation without requiring training sequences in CDMA systems. During the past several years, many researches in this area have focused their effort on the least mean square (LMS) algorithm due to its low complexity. To achieve better performance in suppressing multiple-access interference (MAI) in synchronous CDMA systems, Tang, et al [3]<sup>(1)</sup> applied instead the least mean kurtosis (LMK) algorithm. The LMK algorithm is based on a higher-order statistics rather than the second-order statistics used in the LMS algorithm.

In this paper, a new blind adaptive multiuser detector

-----  
This research was partially supported by New Jersey Center for Wireless Telecommunications.

(1)Note that in [3] only synchronous case was considered.

termed prediction least mean kurtosis (PLMK) algorithm, is proposed for joint MAI and narrowband interference (NBI) suppression in asynchronous CDMA systems. Unlike the regular LMK, it takes into consideration samples earlier than those correspond to current bit. For comparison purposes, we also apply the regular LMK algorithm of [3] to the case of asynchronous CDMA systems. Simulation results show that the blind adaptive multiuser detector with PLMK algorithm provides significantly better performance than the one with LMK algorithm.

## 2. SYSTEM MODEL

We consider the low-pass equivalent model of an asynchronous CDMA system. The received signal due to the  $k$ th user is given by

$$r_k(t) = \sum_{i=-\infty}^{\infty} \sqrt{P_k} b_k s_k(t - iT - \tau_k) \quad (1)$$

where  $T$  is the bit interval,  $b_k \in \{-1, 1\}$  is the information data of the  $k$ th user.  $P_k$  and  $\tau_k$  denote the power and relative delay of the  $k$ th user, respectively. The spreading waveform  $s_k(t)$  is given by

$$s_k(t) = \sum_{n=1}^{N-1} a_k(n) \psi(t - nT_c) \quad (2)$$

where  $a_k(n) \in \{-1, 1\}$  is the  $n$ th element of the spreading sequence for the  $k$ th user,  $N$  is the processing gain and  $T_c = T/N$  is the chip duration.  $\psi(t)$  is a normalized rectangular pulse of width  $T_c$ , i.e.,  $\int_0^{T_c} \psi^2(t) dt = 1$ .

The total received signal can be written as

$$r(t) = \sum_{k=1}^K r_k(t) + i(t) + n(t) \quad (3)$$

where  $K$  is the number of users,  $i(t)$  is the NBI and  $n(t)$  is the white Gaussian noise.

The received signal  $r(t)$  is assumed to pass through a chip-matched filter sampled at chip rate and synchronized to chip time. The  $l$ th received signal sample at the output of the chip-matched filter is

$$r(l) = \int_{\pi_c}^{(l+1)\pi_c} r(t) \psi(t - lT_c) dt \quad (4)$$

from which the  $l$ th NBI sample and the  $l$ th white Gaussian noise sample at the output of the chip-matched filter are  $i(l) = \int_{\pi_c}^{(l+1)\pi_c} i(t) \psi(t - lT_c) dt$  and

$$n(l) = \int_{\pi_c}^{(l+1)\pi_c} n(t) \psi(t - lT_c) dt \text{ respectively.}$$

In this paper, we assume that the NBI is modeled as a  $p$ th-order AR process, i.e.,

$$i(l) = -\sum_{j=1}^p a_j i(l-j) + e(l) \quad (5)$$

where  $e(l)$  is a white Gaussian process with variance  $\varepsilon^2$ .

### 3. BLIND PREDICTION LMK ALGORITHM

Without loss of generality, we assume that the power and the delay of the desired signal are, respectively,  $P_1 = 1$  and  $\tau_1 = 0$ , and convenience, we define  $\tau_k = d_k T_c$  where  $d_k$  is integer between 0 and  $N-1$ . In [3], the LMK algorithm is based on the received signal samples vector  $\mathbf{r}^T = [r(0), r(1), \dots, r(N-1)]$ . It is well known that the current value of NBI is predictable from its past values. Therefore, we expect better performance by extending the received signal samples vector into the interval  $[-MT_c, T]$  ( $M > 0$ ), i.e.,  $\mathbf{r}^T = [r(-M), r(-M+1), \dots, r(-1), r(0), r(1), \dots, r(N-1)]$ , which is termed PLMK algorithm. We consider the case of  $M < N$  in this paper. For a given relative delay vector  $\mathbf{d} = [d_1, \dots, d_K]^T$ , we can obtain from (1)~(4)

$$\mathbf{r} = \sqrt{P_1} (b_1 \mathbf{a}_1 + b'_1 \mathbf{a}'_1) + \sum_{k=2}^K \sqrt{P_k} (b_k \mathbf{a}_k + b'_k \mathbf{a}'_k + b''_k \mathbf{a}''_k) + \mathbf{i} + \mathbf{n} \quad (6)$$

where for  $-M \leq l \leq N-1$  and  $2 \leq k \leq K$

$$\mathbf{a}_1(l) = [a_1(l)] \chi_{(l \geq 0)} \quad (7)$$

$$\mathbf{a}'_1(l) = [a_1(l+N)] \chi_{(l < 0)} \quad (8)$$

$$\mathbf{a}_k(l) = [a_k(l-d_k)] \chi_{(d_k \leq l < N)} \quad (9)$$

$$\mathbf{a}'_k(l) = [a_k(l+N-d_k)] \chi_{(-N+d_k \leq l < d_k)} \quad (10)$$

$$\mathbf{a}''_k(l) = [a_k(l+2N-d_k)] \chi_{(l < -N+d_k)} \quad (11)$$

with  $\chi_A$  is the indicator function for the set  $A$ ,  $b_k$  is the current bit of the  $k$ th user,  $b'_k$  and  $b''_k$  is one bit or two bits earlier than the current bit of the  $k$ th user, respectively.

From (6), we notice having  $3(K-1)+2 = (3K-1)$  vectors  $\{\sqrt{P_1} \mathbf{a}_1, \sqrt{P_1} \mathbf{a}'_1\}$  and  $\{\sqrt{P_k} \mathbf{a}_k, \sqrt{P_k} \mathbf{a}'_k, \sqrt{P_k} \mathbf{a}''_k\}$ ,  $k=2, \dots, K$ . Depending on the relative delays of the multiuser interferers, we have among these,  $L$  ( $2K \leq L \leq 3K-1$ ) non-zero vectors. For the  $L$  non-zero vectors, we write Eqn.(6) in the form

$$\mathbf{r} = \sum_{k=1}^L b_k \mathbf{p}_k + \mathbf{i} + \mathbf{n} \quad (12)$$

where the non-zero vector  $\mathbf{p}_1$  is the desired signal vector  $\sqrt{P_1} \mathbf{a}_1$ , and  $b_1$  is the desired bit. The set of non-zero vectors  $\{\mathbf{p}_2, \dots, \mathbf{p}_L\}$  consists of the intersymbol interference (ISI)  $\{\sqrt{P_1} \mathbf{a}'_1\}$  and the non-zero MAI vectors of the set  $\{\sqrt{P_k} \mathbf{a}_k, \sqrt{P_k} \mathbf{a}'_k, \sqrt{P_k} \mathbf{a}''_k\}$ ,  $k=2, \dots, K$ .  $\{b_2, \dots, b_L\}$  are data coefficients corresponding to the vectors  $\{\mathbf{p}_2, \dots, \mathbf{p}_L\}$ , respectively. For example,  $b_l = b'_k$  if  $\mathbf{p}_l = \sqrt{P_k} \mathbf{a}'_k$ ,  $2 \leq l \leq L$ ,  $1 \leq k \leq K$ .

We use the following cost function of [3] to suppress interference without requiring training sequence:

$$J_B(\mathbf{h}) = 3[E(\mathbf{r}^T \mathbf{h})^2]^3 - E(\mathbf{r}^T \mathbf{h})^4 \quad (13)$$

Taking the gradient with respect to the vector  $\mathbf{h}$ , we have

$$\nabla J_B(\mathbf{h}) = 12E(\mathbf{r}^T \mathbf{h})^2 E(\mathbf{r}^T \mathbf{h}) \mathbf{r} - 4E(\mathbf{r}^T \mathbf{h})^3 \mathbf{r} \quad (14)$$

The mean value  $E(\mathbf{r}^T \mathbf{h})^2$  will be estimated specially by recursive equation

$$G(n) = \beta G(n-1) + (1-\beta) [\mathbf{r}(n)^T \mathbf{h}(n)]^2 \quad (15)$$

with  $0 < \beta < 1$  is forgetting factor.

Using this estimate and the ensemble estimate of  $E(\mathbf{r}^T \mathbf{h})$ ;  $\mathbf{r}(n)^T \mathbf{h}(n)$ , we can get the following equation

$$\tilde{\nabla}_{J_B} [\mathbf{h}(n)] = 4 \left\{ G(n) - [\mathbf{r}(n)^T \mathbf{h}(n)]^2 \right\} \mathbf{r}(n)^T \mathbf{h}(n) \mathbf{r}(n) \quad (16)$$

Then the steepest decent adaptive weight-update algorithm, PLMK algorithm, can be characterized by

$$\mathbf{h}(n+1) = \mathbf{h}(n) - \frac{1}{4} \mu \{ \tilde{\nabla}_{J_B} [\mathbf{h}(n)] \} \quad (17)$$

with  $\tilde{\nabla}_{J_B} [\mathbf{h}(n)]$  from (16) and  $G(n)$  from (15). We can see that training sequence is not needed, the PLMK algorithm is blind.

#### 4. SIMULATION RESULTS

Simulations results carried out to evaluate the performance of the PLMK algorithm is depicted in Fig.1. For comparison, we add to it the results with regular LMK algorithm [3], but for asynchronous case, which can be obtained from PLMK with  $M = 0$ . In this simulation, we use a three-user CDMA system employing Gold Code of length 7. For calculating the averaged SIR at the  $n$ th iteration, we use expression given by [2];

$$\text{SIR}(n) = \frac{\sum_{j=1}^J [\mathbf{h}(n)^T \mathbf{p}_j]^2}{\sum_{j=1}^J \{ \mathbf{h}(n)^T [\mathbf{r}(n) - b_1(n) \mathbf{p}_1] \}^2}$$

with  $J$  is the number of times the simulations are repeated. Each of the other CDMA users has power  $P$  larger than the desired CDMA user power  $P_1 = 1$ . The delay vector is set to  $\mathbf{d} = [0, 1, 3, 6]^T$ . The NBI is modeled as a first-order AR process with  $a_1 = 0.99$  and power of 3dB higher than the desired signal. The white noise power is set to 0.1. We use  $M = 3$ ,  $P = 10$ ,  $\beta = 0.4$ ,  $\mu = 6 \times 10^{-4}$  and  $J = 500$ . From Fig.1, we can easily see that the PLMK algorithm provides significantly better performance than the regular LMK algorithm with almost the same convergence rate.

#### 5. CONCLUSIONS

In this paper, we proposed a new blind adaptive multiuser detector based on prediction least mean kurtosis (PLMK) algorithm for joint suppressing MAI and NBI in asynchronous CDMA systems. For comparison, we also apply the regular LMK algorithm of [3] to the case of asynchronous CDMA systems. Results show that the blind adaptive multiuser detector with PLMK algorithm provides significantly better performance than the one with regular LMK algorithm.

#### 6. REFERENCES

- [1] O. Tanrikulu and A.G. Constantinides, "Least-mean kurtosis: A novel high-order statistics based adaptive filtering algorithm", *IEE Electron. Lett.*, vol.30, pp. 189-190, 1994.
- [2] M. Honig, U. Madhow and S. Verdu, "Blind adaptive multiuser detection", *IEEE Trans. Inform. Theory*, vol. IT-41, No. 4, pp. 944-960, July 1995.
- [3] Z. Tang, Z. Yang and Y. Yao, "Blind multiuser detector based on LMK criterion", *IEE Electron. Lett.*, vol.35, pp. 267-268, 1999.

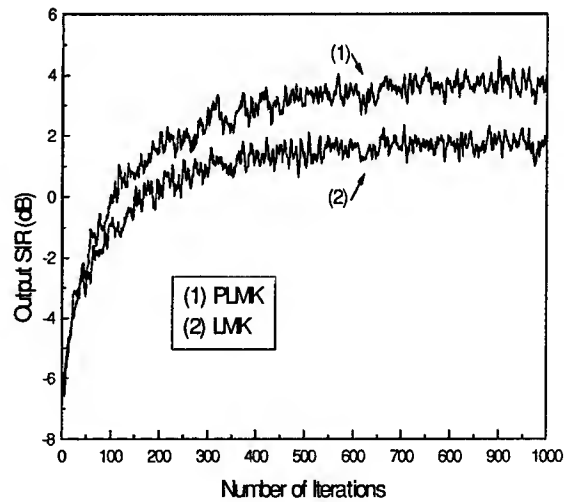


Fig.1 Averaged output SIR versus number of iterations  
(  $N = 7$ ,  $M = 3$ ,  $K = 3$  )

# MMSE EQUALIZATION FOR FORWARD LINK IN 3G CDMA: SYMBOL-LEVEL VERSUS CHIP-LEVEL \*

Thomas P. Krauss, William J. Hillery, and Michael D. Zoltowski

School of Electrical Engineering, Purdue University  
West Lafayette, IN 47907-1285

e-mail: krauss@purdue.edu, hilleryw@ecn.purdue.edu, mikedz@ecn.purdue.edu

## ABSTRACT

We investigate a "symbol-level" MMSE equalizer for the CDMA downlink over a frequency-selective multipath channel meant to improve on the recently proposed "chip-level" downlink equalizers. Indeed the symbol-level equalizer performs better than the chip-level, but is computationally more demanding. The symbol-level equalizer is optimal for "saturated cells" where all Walsh-Hadamard channel codes are in use and have equal power. It performs very close to optimal even for relatively lightly loaded cells. We derive a bound on the off-diagonals of the covariance matrix of the transmitted data that helps explain why the equalizer works when there are fewer active channel codes than the spreading factor. Performance is evaluated through simulations to obtain the average bit error rate (BER) over a class of channels for two cases: no out-of-cell interference, and one equal power base-station. The symbol- and chip-level equalizers are compared to the conventional RAKE receiver.

## 1. INTRODUCTION

Chip-level downlink equalization is a good candidate for improving capacity (in terms of users and/or data rate) in 3G cellular systems such as cdma2000 [1]. These equalizers significantly cancel multi-user access interference (MAI), the main performance limitation for the standard RAKE receiver. The good qualities of the recently proposed "chip-level equalizers" for CDMA downlink are that they need knowledge only of the desired user's spreading code (and long-code), they change only as often as the channel so don't need to be recomputed every symbol, and the same equalizer applies to all users from a given base-station. However, these equalizers do not yield the optimal estimate of the transmitted symbol.

The optimal equalizer is conditioned on all of the channel codes in use and their powers, and also the base-station dependent long code. Since these aren't really random quantities, it should be possible to improve on the performance by using them. One option approaching the optimal one, but still having the nice feature of only needing to know the channel code(s) of the desired user, is derived here. We refer to this as the "symbol-level" equalizer. This equalizer changes every symbol, unlike the chip-level equalizer. We find that this equalizer leads to a performance improvement over the chip-level equalizer when all channel codes are in use and are equal power (in which case the derived equalizer is equal to the optimal symbol estimate). We also make some arguments, and show simulation results, that show this equalizer is applicable when there are fewer active channel

codes per cell.

In this paper we derive the symbol-level MMSE estimator for the two base-station case. One base-station transmits the desired user's data, while the other base-station is considered interference. Spatial diversity and/or oversampling with respect to the chip rate are handled as multiple chip-spaced channels. Our simulations assume spatial diversity is provided by two antennas at the receiver which experience independent fading, and oversample at twice the chip rate.

Some relevant papers on linear chip-level downlink equalizers that restore orthogonality of the Walsh-Hadamard channel codes and hence suppress MAI are [2, 3, 4, 5, 6, 7, 8]. Of these, [4, 7, 8] address antenna arrays, while the others consider a single antenna, possibly with oversampling. In Reference [8] we compare one and two antenna receivers. The interference from other base-stations is addressed in Ghauri and Slock [4], Frank and Visotsky [3], and by Krauss and Zoltowski in [7].

In this paper the channel and noise power are assumed known (i.e., channel estimation error is neglected). Using the exact channel in simulation and analysis leads to an informative upper bound on the performance of these methods, but must be understood as such. For adaptive versions of linear chip equalizers for CDMA downlink see [3] and [6] and some of the references in [5]. [3, 4] present performance analysis in the form of SINR expressions for the multiple base-station case, for the chip-level equalizer. In [7] Krauss and Zoltowski show that the SINR expression along with a Gaussian assumption is a good predictor of uncoded BER for BPSK symbols for the chip-level equalizers.

## 2. DATA AND CHANNEL MODEL

The impulse response for the  $i$ -th antenna channel, between the  $k$ -th base-station transmitter and the mobile-station receiver, is

$$h_i^{(k)}(t) = \sum_{k=0}^{N_a-1} h_i^{(k)}[k] p_{rc}(t - \tau_k) \quad i = 1, 2, k = 1, 2 \quad (1)$$

$p_{rc}(t)$  is the composite chip waveform (including both the transmit and receive low-pass filters) which we assume has a raised-cosine spectrum.  $N_a$  is the total number of delayed paths or "multipath arrivals," some of which may have zero or negligible power without loss of generality.

The channel we consider for this work consists of  $N_a = 17$  equally spaced paths  $0.625\mu s$  apart ( $\tau_0 = 0$ ,  $\tau_1 = 0.625\mu s$ , ...); this yields a delay spread of at most  $10\mu s$ , which is an upper bound for most channels encountered in urban cellular systems. We model the class of channels with 4 equal-power random coefficients with arrival times picked randomly from the set  $\{\tau_0, \tau_1, \dots, \tau_{16}\}$ ; the rest of the coefficients  $h_i^{(1)}[k]$  are zero. For base-station 1, once the 4 arrival times have

\*THIS RESEARCH WAS SUPPORTED BY THE TEXAS INSTRUMENTS DSP UNIVERSITY RESEARCH PROGRAM AND THE AIR FORCE OFFICE OF SCIENTIFIC RESEARCH UNDER GRANT NO. F49620-00-1-0127.

been picked at random and then sorted, the first and last arrival times are forced to be at 0 and the maximum delay spread of  $10\mu s$  respectively. Base-station 2's arrival times are chosen in the same fashion and independent of base-station 1's, but *without* forcing arrivals at 0 and  $10\mu s$ . The coefficients are equal-power, complex-normal random variables, independent of each other. The arrival times at antennas 1 and 2 associated with a given base-station are the same, but the coefficients are independent.

The "multi-user chip symbols" for base-station  $k$ ,  $s^{(k)}[n]$ , may be described as

$$s^{(k)}[n] = c_{bs}^{(k)}[n] \sum_{j=1}^{N_u^{(k)}} \sum_{m=0}^{N_s-1} \alpha_j^{(k)} b_j^{(k)}[m] c_j^{(k)}[n - N_c m] \quad (2)$$

where the various quantities are defined as follows:  $c_{bs}^{(k)}[n]$  is the base-station dependent long code;  $\alpha_j^{(k)}$  is the  $j^{th}$  user's gain;  $b_j^{(k)}[m]$  is the  $j^{th}$  user's bit/symbol sequence;  $c_j^{(k)}[n]$ ,  $n = 0, 1, \dots, N_c - 1$ , is the  $j^{th}$  user's channel (short) code;  $N_c$  is the length of each channel code (assumed the same for each user);  $N_u^{(k)}$  is the total number of active users;  $N_s$  is the number of bit/symbols transmitted during a given time window. The signal received at the  $i^{th}$  antenna (after convolving with a matched filter impulse response having a square-root raised cosine spectrum) from base-station  $k$  is

$$y_i^{(k)}(t) = \sum_n s^{(k)}[n] h_i^{(k)}(t - nT_c) \quad i = 1, 2 \quad (3)$$

where  $h_i^{(k)}(t)$  is as defined in Eqn. (1). The total received signal at the mobile-station is simply the sum of the contributions from the different base-stations plus noise:

$$y_i(t) = y_i^{(1)}(t) + y_i^{(2)}(t) + \eta_i(t) \quad i = 1, 2. \quad (4)$$

$\eta_i(t)$  is a noise process assumed white and gaussian prior to coloration by the receiver chip-pulse matched filter.

For the first antenna, we oversample the signal  $y_1(t)$  in Eqn. (4) at twice the chip-rate to obtain  $y_1[n] = y_1(nT_c)$  and  $y_2[n] = y_1(\frac{T_c}{2} + nT_c)$ . These discrete-time signals have corresponding impulse responses  $h_1^{(k)}[n] = h_1^{(k)}(t)|_{t=nT_c}$  and  $h_2^{(k)}[n] = h_1^{(k)}(t)|_{t=\frac{T_c}{2}+nT_c}$  for base-stations  $k = 1, 2$ .

For the second antenna, we also oversample the signal  $y_2(t)$  in Eqn. (4) at twice the chip-rate to obtain  $y_3[n] = y_2(nT_c)$  and  $y_4[n] = y_2(\frac{T_c}{2} + nT_c)$ . These discrete-time signals have corresponding impulse responses  $h_3^{(k)}[n] = h_2^{(k)}(t)|_{t=nT_c}$  and  $h_4^{(k)}[n] = h_2^{(k)}(t)|_{t=\frac{T_c}{2}+nT_c}$  for base-stations  $k = 1, 2$ .

Let  $M$  denote the total number of chip-spaced channels due to both receiver antenna diversity and / or oversampling.

### 3. CHIP-LEVEL EQUALIZER

The "Chip-level" MMSE equalizer is shown in Figure 1 (two antenna case with no oversampling). It estimates the multi-user synchronous sum signal for either base-station 1 or 2, and then correlates with the desired user's channel code times that base-station's long code. To derive the chip-level MMSE equalizer, it is useful to define signal vectors and channel matrices based on the equalizer length  $N_g$ . The "recovered" chip signal will be  $\hat{s}^{(k)}[n - D] = \mathbf{g}^{(k)H} \mathbf{y}[n]$  for some delay  $D$ , where  $\mathbf{g}^{(k)}$  is the  $MN_g \times 1$  chip-level equalizer for

base-station  $k$ ,  $k = 1, 2$ . The equalizer coefficients  $g_i^{(k)}[n]$  comprise the equalizer vector

$$\mathbf{g}^{(k)} = [\mathbf{g}_1^{(k)T} \dots \mathbf{g}_M^{(k)T}]^T \quad (5)$$

where

$$\mathbf{g}_i^{(k)} = [g_i^{(k)}[0], g_i^{(k)}[1], \dots, g_i^{(k)}[N_g - 1]]^T \quad i = 1, \dots, M. \quad (6)$$

The  $MN_g \times 1$  vectorized received signal is given by

$$\mathbf{y}[n] = \mathbf{H}^{(1)} \mathbf{s}^{(1)}[n] + \mathbf{H}^{(2)} \mathbf{s}^{(2)}[n] + \boldsymbol{\eta}[n] \quad (7)$$

where

$$\mathbf{s}^{(k)}[n] = [s^{(k)}[n], s^{(k)}[n-1], \dots, s^{(k)}[n - (N_g + L - 2)]]^T \quad (8)$$

$$\mathbf{H}^{(k)} = \begin{bmatrix} \mathbf{H}_1^{(k)} \\ \vdots \\ \mathbf{H}_M^{(k)} \end{bmatrix} \quad (9)$$

$\mathbf{H}_i^{(k)}$  is the  $N_g \times (L + N_g - 1)$  convolution matrix

$$\mathbf{H}_i^{(k)} = \begin{bmatrix} h_i^{(k)}[0] & 0 & \dots & 0 \\ h_i^{(k)}[1] & h_i^{(k)}[0] & 0 & 0 \\ \vdots & \ddots & \ddots & \vdots \\ h_i^{(k)}[L-1] & h_i^{(k)}[L-2] & \ddots & h_i^{(k)}[0] \\ 0 & h_i^{(k)}[L-1] & \ddots & h_i^{(k)}[1] \\ \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & 0 & h_i^{(k)}[L-1] \end{bmatrix}^T. \quad (10)$$

Equation (7) is more compactly written as

$$\mathbf{y}[n] = \mathcal{H} \mathbf{s}[n] + \boldsymbol{\eta}[n] \quad (11)$$

where

$$\mathcal{H} = [\mathbf{H}^{(1)} \vdots \mathbf{H}^{(2)}] \quad (12)$$

and

$$\mathbf{s}[n] = [\mathbf{s}^{(1)T}[n] \quad \mathbf{s}^{(2)T}[n]]^T. \quad (13)$$

The MMSE criterion is

$$\min_{\mathbf{g}^{(k)}} E\{\|\mathbf{g}^{(k)H}(\mathcal{H} \mathbf{s}[n] + \boldsymbol{\eta}[n]) - \boldsymbol{\delta}_D^T \mathbf{s}^{(k)}[n]\|^2\} \quad (14)$$

where  $\boldsymbol{\delta}_D$  is all zeroes except for unity in the  $(D + 1) - th$  position (so that  $\boldsymbol{\delta}_D^T \mathbf{s}^{(k)}[n] = s^{(k)}[n - D]$ ).

We assume unit energy signals,  $E\{|s^{(k)}[n]|^2\} = 1$ , and furthermore that the chip-level symbols  $s^{(k)}[n]$  are independent and identically distributed,  $E\{\mathbf{s}[n] \mathbf{s}^H[n]\} = \mathbf{I}$ . This is the case if the base-station dependent long codes,  $c_{bs}^{(k)}[n]$ , are treated as iid sequences, a very good assumption in practice. The equalizer  $\mathbf{g}^{(k)}$  which attains the minimum is

$$\mathbf{g}^{(k)} = (\mathcal{H} \mathcal{H}^H + \mathbf{R}_{\eta\eta})^{-1} \mathbf{H}^{(k)} \boldsymbol{\delta}_D. \quad (15)$$

The MMSE is

$$\text{MMSE} = 1 - \boldsymbol{\delta}_D^T \mathbf{H}^{(k)H} (\mathcal{H} \mathcal{H}^H + \mathbf{R}_{\eta\eta})^{-1} \mathbf{H}^{(k)} \boldsymbol{\delta}_D. \quad (16)$$

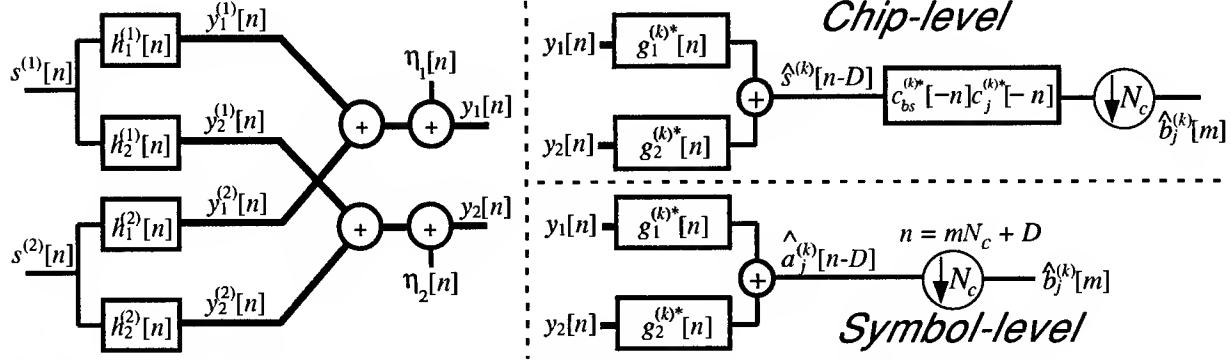


Figure 1. Chip and Symbol MMSE Estimators for  $k^{\text{th}}$  Base-Station, two antennas, no oversampling.

The MMSE equalizer is a function of the delay  $D$ . The MMSE may be computed for each  $D$ ,  $0 \leq D \leq N_g + L - 2$  with only one matrix inversion (which has to be done to form  $\mathbf{g}^{(k)}$  anyway). Once the  $D$  yielding the smallest MMSE is determined, the corresponding equalizer  $\mathbf{g}^{(k)}$  may be computed without further matrix inversion or system solving.

#### 4. SYMBOL-LEVEL EQUALIZER

In this section we present what we call the “symbol-level” MMSE estimator. This estimator depends on the user index and symbol index, and hence varies from symbol to symbol. The FIR estimator that we derive here is a simplified version of that presented in [9] where in our case, all the channels and delays from a given base-station are the same. The conclusions reached in that paper apply equally well here, namely that FIR MMSE equalization always performs at least as well as the “coherent combiner” (that is, the RAKE receiver). This type of symbol-level receiver has also been presented in [10], although again not specifically for the CDMA downlink.

The symbol-level equalizer differs from the chip-level equalizer in that the base station and Walsh-Hadamard codes do not appear explicitly in the block diagram (see Figure 1). Instead, the codes become incorporated into the equalizer itself. To derive the equalizer, we first define  $a_j^{(k)}[n]$  as the bit sequence  $b_j^{(k)}[m]$  upsampled by  $N_c$ :  $a_j^{(k)}[n] = b_j^{(k)}[m]$  when  $n = mN_c$  and  $a_j^{(k)}[n] = 0$  otherwise. We wish to estimate  $b_j^{(k)}[m]$  directly and we do this by finding

$$\min_{\mathbf{g}^{(k)}} E\{|\hat{a}_j^{(k)}[n-D] - a_j^{(k)}[n-D]|^2\} \quad (17)$$

where the minimization is done only when  $n-D = mN_c$ . As in the chip-level case,  $\hat{a}_j^{(k)}[n-D] = \mathbf{g}^{(k)H} \mathbf{y}[n]$  where  $\mathbf{y}[n]$  is given by Eq. (11). Setting  $n = mN_c + D$ , the MSE is minimized yielding

$$\mathbf{g}^{(k)}[m] = (\mathcal{H} \mathbf{R}_{SS}[mN_c + D] \mathcal{H}^H + \mathbf{R}_{\eta\eta})^{-1} \mathbf{R}_{bS}[m] \quad (18)$$

where

$$\mathbf{R}_{SS}[n] = E\{\mathbf{s}[n] \mathbf{s}^H[n]\} \quad (19)$$

$$\mathbf{R}_{bS}[m] = E\{b_j^*[m] \mathbf{s}[mN_c + D]\} \quad (20)$$

We now proceed to derive expressions for  $\mathbf{R}_{SS}[n]$  and  $\mathbf{R}_{bS}[m]$ . Using Eq. (13),

$$\mathbf{R}_{SS}[n] = \begin{bmatrix} \mathbf{R}_{SS}^{(11)}[n] & \mathbf{R}_{SS}^{(12)}[n] \\ \mathbf{R}_{SS}^{(21)}[n] & \mathbf{R}_{SS}^{(22)}[n] \end{bmatrix} \quad (21)$$

where  $\mathbf{R}_{SS}^{(pq)}[n] = E\{\mathbf{s}^{(p)}[n] \mathbf{s}^{(q)H}[n]\}$ . We assume here that the desired user is only transmitted by base station  $k$ . We also assume that the base station and Walsh-Hadamard codes are deterministic and known so that the only random elements in  $\mathbf{s}[n]$  are the transmitted bits. Then  $E\{s^{(k)}[n] s^{(j)*}[m]\} = 0$  for  $k \neq j$  and any  $n$  and  $m$ , so  $\mathbf{R}_{SS}^{(12)}[n] = \mathbf{R}_{SS}^{(21)}[n] = \mathbf{0}$ . The  $(i, j)^{\text{th}}$  element of  $\mathbf{R}_{SS}^{(kk)}[n]$  is  $S_{ij}^{(kk)}[n] = E\{s^{(k)}[n+1-i] s^{(k)*}[n+1-j]\}$ . When  $i = j$ ,  $S_{ij}^{(kk)}[n] = 1$ . When  $i \neq j$ ,

$$S_{ij}^{(kk)}[n] = \begin{cases} \frac{1}{2N_u^{(k)}} B_{ij}[n] W_{ij}^{(k)}[n], & \text{when} \\ & (n+1-i) \bmod N_c = (n+1-j) \bmod N_c \\ 0 & \text{otherwise} \end{cases} \quad (22)$$

where

$$B_{ij}[n] = c_{bs}^{(k)}[n+1-i] c_{bs}^{(k)*}[n+1-j] \in \{\pm 2, \pm 2j\} \forall i, j \quad (23)$$

$$W_{ij}^{(k)}[n] = \sum_{p=1}^{N_u^{(k)}} c_p^{(k)}[(n+1-i) \bmod N_c] c_p^{(k)*}[(n+1-j) \bmod N_c] \quad (24)$$

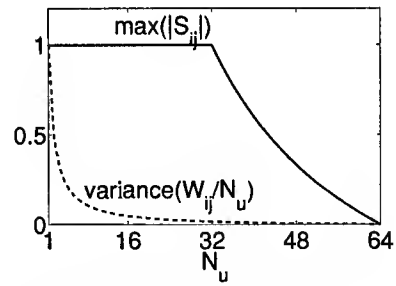


Figure 2. Bound on the potentially non-zero off-diagonal elements of  $\mathbf{R}_{SS}[n]$  [ $N_c = 64$ ].

With fixed  $m$  and  $n$ ,  $m \neq n$ , note that  $[c_1^{(k)}[m], \dots, c_{N_c}^{(k)}[m]]^T$  and  $[c_1^{(k)}[n], \dots, c_{N_c}^{(k)}[n]]^T$  are two different rows of the Hadamard matrix. The element-by-element (Schur) product of these two rows is also a row of the Hadamard matrix containing  $(N_c/2)$  1's and  $(N_c/2)$  -1's. So

$$|W_{ij}^{(k)}[n]| \leq \begin{cases} N_u & N_u = 1, \dots, N_c/2 \\ N_c - N_u & N_u = N_c/2 + 1, \dots, N_c \end{cases} \quad (25)$$



Therefore, when  $i \neq j$  and  $(n+1-i) \bmod N_c = (n+1-j) \bmod N_c$ ,

$$|S_{ij}^{(kk)}[n]| \leq \begin{cases} 1 & N_u^{(k)} = 1, \dots, N_c/2 \\ N_c/N_u^{(k)} - 1 & N_u^{(k)} = N_c/2 + 1, \dots, N_c \end{cases} \quad (26)$$

This bound is plotted as a function of  $N_u^{(k)}$  in Fig. 2. Note that when  $N_u^{(k)} = N_c$ ,  $S_{ij}^{(kk)}[n] = 0$  for all  $i \neq j$ , so  $\mathbf{R}_{SS}^{(kk)}[n] = \mathbf{I}$ . If we assume that the  $N_u^{(k)}$  Walsh codes are chosen randomly when  $N_u^{(k)} < N_c$ , it can be shown that  $W_{ij}^{(k)}[n]/N_u^{(k)}$  is a linear function of a hypergeometric random variable. Its variance is  $N_u^{(k)}(N_c - N_u^{(k)})/(N_c - 1)$ . Therefore, those off-diagonal elements which are not zero have zero mean and the variance shown in the plot in Fig. 2. For nearly all values of  $N_u^{(k)}$ , the variance is clearly quite small. So in all cases, we may well approximate  $\mathbf{R}_{SS}^{(kk)}[n]$  by  $\mathbf{I}$  in Eq. (18) yielding

$$\mathbf{g}^{(k)}[m] = (\mathcal{H}\mathcal{H}^H + \mathbf{R}_{\eta\eta})^{-1} \mathcal{H}\mathbf{R}_{bS}[m] \quad (27)$$

We will see through simulation that this approximation works quite well when compared to the "exact" equalizer constructed with a time-varying  $\mathbf{R}_{SS}$ .

The  $i^{\text{th}}$  element of  $\mathbf{R}_{bS}[m]$  is (with  $n = mN_c + D$ ):

$$E\{b_j^{(k)*}[m]s[n+1-i]\} = \begin{cases} c_{bs}^{(k)}[n+1-i]c_j^{(k)}[D+1-i], & \text{for } 0 \leq D+1-i \leq N_c-1 \\ 0 & \text{otherwise} \end{cases} \quad (28)$$

With  $D$  satisfying  $N_c - 1 \leq D \leq L + N_g - 2$ , the entire Walsh code for the desired user appears in  $\mathbf{R}_{bS}[m]$  and

$$\mathbf{R}_{bS}[m] = [\mathbf{0}_{D+1-N_c} \quad \mathbf{c}_j[m] \quad \mathbf{0}_{L+N_g-2-D}]^T \quad (29)$$

where

$$\mathbf{c}_j[m] = [c_{bs}^{(k)}[(m+1)N_c-1]c_j^{(k)}[N_c-1], \dots, c_{bs}^{(k)}[mN_c+1]c_j^{(k)}[1], c_{bs}^{(k)}[mN_c]c_j^{(k)}[0]]^T \quad (30)$$

While the equalizer varies from symbol to symbol due to variation in both  $\mathbf{R}_{SS}[n]$  and  $\mathbf{R}_{bS}[m]$ , by approximating  $\mathbf{R}_{SS}[n]$  by  $\mathbf{I}$ , the variation is confined to  $\mathbf{R}_{bS}[m]$ .

## 5. RAKE RECEIVER

The RAKE receiver is simply a multipath-incorporating matched filter. In particular, the RAKE can be viewed as a chip-spaced filter matched to the channel, followed by correlation with the long code times channel code. Note, in practice, these operations are normally reversed, but may be reversed due to short-time LTI assumptions. The RAKE receiver is exactly represented by the "Chip-Level" portion of Figure 1, if we let  $N_g = L$  and  $g_i^{(k)}[n] = h_i^{(k)}[L-n]$ ,  $n = 0, \dots, L-1$ ,  $i = 1, \dots, M$ .

## 6. SIMULATION RESULTS

A wideband CDMA forward link was simulated similar to one of the options in the US cdma2000 proposal [1]. The spreading factor is  $N_c = 64$  chips per bit. Simulations were performed for both "saturated cells," that is, all 64 possible channel codes active, as well as lightly loaded cells with 8 channel codes active. The chip rate is 3.6864 MHz ( $T_c = 0.27\mu s$ ), 3 times that of IS-95. The data symbols are BPSK which, for each user, are spread with a length 64

Walsh-Hadamard sequence. The signals for all the users are of equal power and summed synchronously, and each base-station had the same number of users. The sum signal is scrambled with a multiplicative QPSK spreading sequence ("scrambling code") of length 32768 similar to the IS-95 standard.

The uncoded BER results are averaged over different channels for varying SNRs. The channels were generated according to the model presented in Section 2. "SNR" is defined to be the ratio of the sum of the average powers of the received signals from the desired base-station, to the average noise power, after chip-matched filtering. "SNR per user per symbol" is the SNR divided by the number of users and multiplied by the spreading factor. For the chip-level MMSE, the total delay of the signal,  $D$ , through both channel and equalizer, was chosen to minimize the MSE of the equalizer.

We first present results for a receiver near the base-station so that out-of-cell interference is negligible. Two receive antennas are employed with no oversampling. Two equalizer lengths were simulated: for chip-level,  $N_g = 57$  and 114, while for symbol-level, the length is chosen  $N_c - 1$  longer. Since the chip-level equalizer is followed by correlation with the channel code times long code, its effective length is  $N_g + N_c - 1$ ; hence, a fair comparison between the symbol-level and chip-level sets the symbol-level equalizer longer by  $N_c - 1$  chips. Figure 3 presents the results for the fully loaded cell case, i.e. 64 equal power users were simulated. The RAKE receiver is significantly degraded at high SNR by the MAI, which is seen in the Figure as a BER floor for SNR greater than 10 dB. The chip- and symbol-level equalizers perform much better than the RAKE. Increasing the equalizer length improves performance for both chip-level and symbol-level. Comparing the length 57 chip-level to 120 symbol-level, we observe little improvement in the symbol level at low SNR with increasing improvement, up to 2-3 dB, at high SNR. Comparing length 114 chip-level to 177 symbol-level also shows an improvement that increases with SNR, but less of an improvement than for the shorter equalizers. Note that since all 64 channel codes are present and have equal power,  $\mathbf{R}_{ss} = \mathbf{I}$  and the symbol-level MMSE estimate is optimal in the MSE sense.

In Figure 4, once again the out-of-cell interference is assumed negligible. In this simulation only 8 equal power channel codes are active, i.e., the cell is only lightly to moderately loaded. In this simulation the RAKE receiver does much better since it experiences less in-cell MAI than for 64 users. For the range of SNR simulated the chip-level equalizer does only slightly better than the RAKE receiver. As for the fully loaded cell, the symbol-level equalizer performs better than the chip-level equalizer. For comparison the "optimal" symbol-level equalizer is shown which involves a matrix inverse for every symbol (as in Equation (18)); this equalizer is only slightly better than the symbol-level equalizer presented in this paper. This result justifies the assumption / simplification that  $\mathbf{R}_{ss}$  is proportional to  $\mathbf{I}$ , even when  $N_u < N_c$ .

Figure 5 results from a simulation with two base-stations, each with 64 equal power users. The 2<sup>nd</sup> base-station is treated as interference and is received with the same power as the 1<sup>st</sup>, desired user's base-station. Specifically,

$$\sum_{m=1}^M E\{|y_m^{(1)}[n]|^2\} = \sum_{m=1}^M E\{|y_m^{(2)}[n]|^2\}. \quad (31)$$

In addition to two independent antennas, two-times oversampling is employed for a total of four chip-spaced channels. The results are very analogous to the single base-station case: the symbol-level outperforms the chip-level, increasingly so at high SNR. However the improvement is more dramatic, especially for the shorter lengths.

## 7. CONCLUSIONS

The symbol-level equalizer derived here performs better than the chip-level, however at a greater computational cost. In fact our simulations have shown that even though the equalizer is sub-optimal, it has performance closely approaching optimality. The approximation that the source covariance is diagonal means that a matrix inverse is required only as often as the channel changes (and not every symbol), and hence the computational complexity is much smaller than the optimal equalizer.

## REFERENCES

- [1] Telecommunications Industry Association, "Physical Layer Standard for cdma2000 Standards for Spread Spectrum Systems - T1A/E1A/1S-2000.2-A", T1A/E1A Interim Standard, March 2000.
- [2] Anja Klein, "Data Detection Algorithms Specially Designed for the Downlink of CDMA Mobile Radio Systems", in *IEEE 47th Vehicular Technology Conference Proceedings*, pp. 203-207, Phoenix, AZ, May 4-7 1997.
- [3] Colin D. Frank and Eugene Visotsky, "Adaptive Interference Suppression for Direct-Sequence CDMA Systems with Long Spreading Codes", in *Proceedings 36th Allerton Conf. on Communication, Control, and Computing*, pp. 411-420, Monticello, IL, Sept. 23-25 1998.
- [4] I. Ghauri and DTM. Slock, "Linear receivers for the DS-SS-CDMA downlink exploiting orthogonality of spreading sequences", in *Conf. Rec. 32nd Asilomar Conf. on Signals, Systems, and Computers, Pacific Grove, CA*, Nov. 1998.
- [5] Kari Hooli, Matti Latva-aho, and Markku Juntti, "Multiple Access Interference Suppression with Linear Chip Equalizers in WCDMA Downlink Receivers", in *Proc. Global Telecommunications Conf.*, pp. 467-471, Rio de Janeiro, Brazil, Dec. 5-9 1999.
- [6] Stefan Werner and Jorma Lilleberg, "Downlink Channel Decorrelation in CDMA Systems with Long Codes", in *IEEE 49th Vehicular Technology Conference Proceedings*, vol. 2, pp. 1614-1617, Houston, TX, May 16-19 1999.
- [7] Thomas P. Krauss and Michael D. Zoltowski, "MMSE Equalization Under Conditions of Soft Hand-Off", in *IEEE Sixth International Symposium on Spread Spectrum Techniques & Applications (ISSSTA 2000) (to appear)*, September 6-8 2000.
- [8] T. Krauss and M. Zoltowski, "Oversampling Diversity Versus Dual Antenna Diversity for Chip-Level Equalization on CDMA Downlink", in *Proceedings of First IEEE Sensor Array and Multichannel Signal Processing Workshop*, Cambridge, MA, March 16-17 2000.
- [9] Hui Liu and Mike Zoltowski, "Blind equalization in antenna array CDMA systems", *IEEE Transactions on Signal Processing*, vol. 45, pp. 161-172, Jan. 1997.
- [10] A. Klein, G. Kaleh, and P. Baier, "Zero Forcing and Minimum Mean-Square-Error Equalization for Multiuser Detection in Code-Division Multiple-Access Channels", *IEEE Transactions on Vehicular Technology*, vol. 45, pp. 276-287, May 1996.

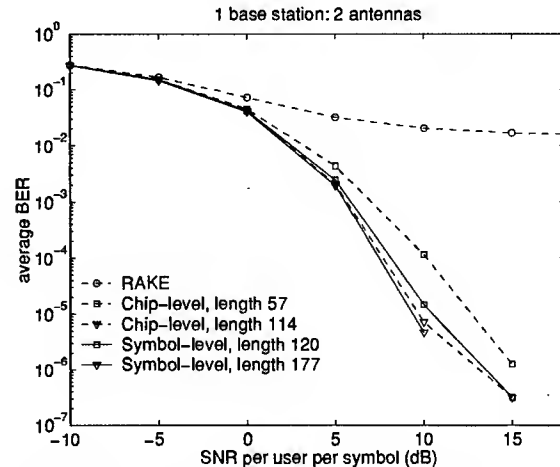


Figure 3. Fully loaded cell, all 64 channel codes in use.

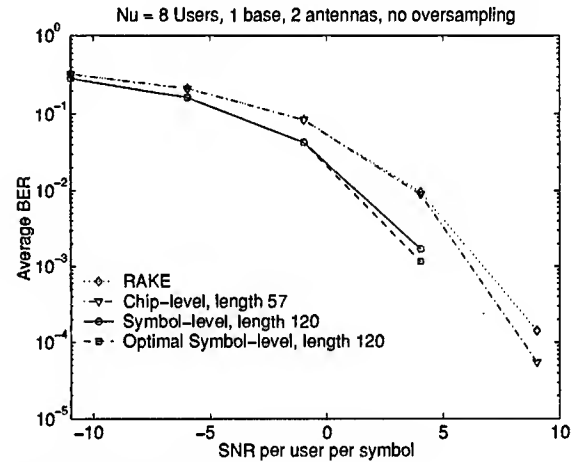


Figure 4. Lightly loaded cell, 8 out of 64 active channel codes.

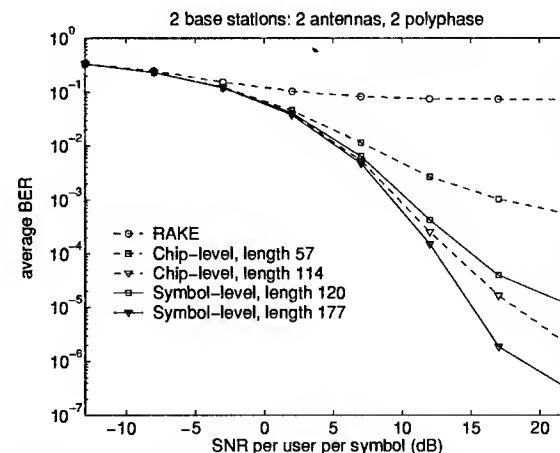


Figure 5. One interfering base-station of equal power, 64 channel codes per cell.

# TRANSFORM DOMAIN ARRAY PROCESSING FOR CDMA SYSTEMS

Yimin Zhang<sup>†</sup>, Kehu Yang<sup>‡</sup> and Moeness G. Amin<sup>†</sup>

<sup>†</sup> Department of Electrical and Computer Engineering,  
Villanova University, Villanova, PA 19085  
E-mail: zhang@ece.vill.edu, moeness@ece.vill.edu

<sup>‡</sup> ATR Adaptive Communications Research Laboratories,  
Seika-cho, Soraku-gun, Kyoto 619-0288, Japan  
E-mail: yang@acr.atr.co.jp

## ABSTRACT

*In this paper, we propose transform domain array processing schemes for DS-CDMA communications. Space-time adaptive processing (STAP) is a useful means to combat the multiuser interference (MUI) in CDMA systems. The computation burden and slow convergence are two major problems in implementing the STAP. This paper proposes optimum and sub-optimum transform domain arrays with different feedback schemes for CDMA communications. The transform domain arrays provide reduced computations over traditional implementation methods as well as they offer improved convergence performance, leading to an efficient system implementation.*

## 1. INTRODUCTION

Array processing in direct-sequence code division multiple access (DS-CDMA) communications has recently attracted considerable attention [1, 2, 3]. The use of the joint space-time adaptive processing (STAP), which includes two-dimensional RAKE (2-D RAKE) receiver, provides excellent performance of suppressing the multiuser interference (MUI) and inter-symbol interference (ISI) as well as combining the multipath signals to achieve the RAKE diversity effect in frequency-selective fading. In order to combine sufficient number of multipath rays to enhance the signal power and reduce the ISI, a large number of weights are required at the feedback loop. The complexity and convergence rate problems remain the bottleneck of the implementation of these systems [4].

In this paper, we propose a transform domain approach to chip-level space-time adaptive processing for DS-CDMA communications with different feedback schemes. Chip-level space-time adaptive processing effectively mitigates both MUI and ISI before despreading and, as such, only a simple correlation and summation operation with the desired user's code is required to follow. When subband array is applied to the chip-rate STAP processing, the signal decorrelation using orthogonal transforms and feedback schemes greatly reduce the circuit size within each single

feedback loop, and subsequently improves the receiver convergence performance [5, 6]. Discrete Fourier Transform (DFT), filter banks and wavelets are among the commonly used orthogonal transform for this purpose [7]. In this paper, we consider the DFT as the example. Decimation available at the transform domain processing also makes it possible to reduce the signal processing speed at each transform domain bin [5, 6].

## 2. SPACE-TIME ADAPTIVE PROCESSING FOR CDMA

We consider a base station using an antenna array of  $N$  sensors with  $P$  users. In CDMA systems, usually  $P > N$ . The received signal vector at the array is expressed, in discrete-time form sampled at the chip rate, as

$$\underline{x}(k) = \sum_{p=1}^P \sum_{l=-\infty}^{\infty} d_p(l) \underline{h}_p(k-l) + \underline{b}(k) \quad (1)$$

where  $d_p(k)$  and  $\underline{h}_p(k)$  are the chip-rate sequence and the channel response vector of the  $p$ th user, and  $\underline{b}(k)$  is the additive noise vector.

In CDMA communications, each symbol is spread into  $L$  chips. Without loss of generality, we denote the signal of the user of interest as  $s_1(n)$ , and the signals from other users as  $s_p(n)$ ,  $p = 2, \dots, P$ . Aperiodic spreading sequence are assumed. The chip length is  $L = T/T_c$ , where  $T$  and  $T_c$  are, respectively, the symbol duration and chip duration. We denote the spreading sequence for the  $n$ th symbol of the  $P$  users as  $c_p(n, l)$ ,  $p = 1, \dots, P$ ,  $l = 1, \dots, L$ . Then,

$$d_p(k) = s_p(n) c_p(n, l - l_p) \quad (2)$$

where  $k = nL + l$ , and  $l_p$  ( $0 \leq l_p < L$ ) is the chip delay index that models the asynchronous system. We make the following assumptions:

A1) The information symbols  $s_p(n)$ ,  $p = 1, 2, \dots, P$ , are wide-sense stationary and i. i. d. with  $E[s_p(n)s_p^*(n)] = 1$ .

A2) The spreading sequences  $c_p(n, l)$ ,  $p = 1, 2, \dots, P$ ,  $l = 1, \dots, L$ , are assumed independent random sequences.

The work of Y. Zhang and M. G. Amin is supported by the Office of Naval Research under Grant N00014-98-1-0176.

A3) All channels  $\mathbf{h}_p(k)$ ,  $p = 1, 2, \dots, P$ , are linear time-invariant, and of a finite duration within  $[0, DT_c]$ . That is,  $\mathbf{h}_p(k) = 0$ ,  $p = 1, 2, \dots, P$ , for  $k > D$  and  $k < 0$ .

A4) The noise vector  $\mathbf{b}(k)$  is zero-mean, temporally and spatially white with

$$E[\mathbf{b}(k)\mathbf{b}^T(k+l)] = \mathbf{0} \text{ for any } l$$

and

$$E[\mathbf{b}(k)\mathbf{b}^H(k+l)] = \sigma \mathbf{I}_N \delta(l),$$

where the superscripts  $T$  and  $H$  denote transpose and conjugate transpose, respectively,  $\mathbf{I}_N$  is the  $N \times N$  identity matrix, and  $\delta(l)$  is the Kronecker delta function.

By stacking  $M$  consecutive chips of  $\mathbf{x}(k)$ , we can obtain

$$\mathbf{x}(k) = \sum_{p=1}^P \mathcal{H}_p \mathbf{d}_p(k) + \mathbf{b}(k) = \mathcal{H} \mathbf{d}(k) + \mathbf{b}(k), \quad (3)$$

where

$$\mathbf{x}(k) = [\mathbf{x}^T(k) \mathbf{x}^T(k-1) \dots \mathbf{x}^T(k-M+1)]^T, \quad (4)$$

$$\mathbf{d}_p(k) = [d_p(k) d_p(k-1) \dots d_p(k-M+1)]^T, \quad (5)$$

$$\mathbf{d}(k) = [\mathbf{d}_1^T(k) \mathbf{d}_2^T(k) \dots \mathbf{d}_P^T(k)]^T, \quad (6)$$

$$\mathcal{H}_p = \begin{bmatrix} \mathbf{h}_p(0) & \dots & \mathbf{h}_p(D_p) & 0 & \dots & \dots & 0 \\ 0 & \mathbf{h}_p(0) & \dots & \mathbf{h}_p(D_p) & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & \dots & 0 & \mathbf{h}_p(0) & \dots & \mathbf{h}_p(D_p) \end{bmatrix}, \quad (7)$$

$$\mathcal{H} = [\mathcal{H}_1 \mathcal{H}_2 \dots \mathcal{H}_P]^T, \quad (8)$$

and

$$\mathbf{b}(k) = [\mathbf{b}^T(k) \mathbf{b}^T(k-1) \dots \mathbf{b}^T(k-M+1)]^T. \quad (9)$$

Denote  $\mathbf{w}$  as the weight vectors of the STAP system corresponding to  $\mathbf{x}(k)$ , the output of the STAP becomes

$$y(k) = \mathbf{w}^T \mathbf{x}(k). \quad (10)$$

The optimum weight vector under the minimum mean square error (MMSE) criterion

$$\min_{\mathbf{w}} E |y(k) - d_1(k-v)|^2 \quad (11)$$

is given by the Wiener-Hopf solution

$$\mathbf{w}_{opt} = \mathbf{R}^{-1} \mathbf{r}, \quad (12)$$

where  $v \geq 0$  is a delay to minimize the MMSE,

$$\mathbf{R} = E[\mathbf{x}^*(k)\mathbf{x}^T(k)], \quad (13)$$

$$\mathbf{r} = E[\mathbf{x}^*(k)d_1(k-v)], \quad (14)$$

and the superscript  $*$  denotes complex conjugate. The training signal is assumed to be an ideal replica of  $d_1(k)$ .

From the assumptions A1) – A4), (13) and (14) can be expressed as

$$\mathbf{R} = \mathcal{H}^* \mathcal{H}^T + \sigma \mathbf{I}_{MN}, \quad (15)$$

and

$$\mathbf{r} = \mathcal{H}_1^* \mathbf{e}_v, \quad (16)$$

respectively, where

$$\mathbf{e}_v = \underbrace{[0 \dots 0]_v}_{v} [1 \ 0 \dots 0]^T. \quad (17)$$

The MMSE is given by

$$\text{MMSE} = E |\mathbf{w}_{opt}^T \mathbf{x}(k) - d_1(k-v)|^2 = 1 - \mathbf{r}^H \mathbf{R}^{-1} \mathbf{r}. \quad (18)$$

Despreading the array output signal  $\mathbf{y}(k)$  by the signature code of desired signal, we obtain the symbol-rate output signal for detection, expressed as

$$z(n) = \sum_{l=0}^{L-1} y(nL + l + v) c_1(n, l). \quad (19)$$

### 3. TRANSFORM DOMAIN ARRAYS WITH DIFFERENT FEEDBACK SCHEMES

#### 3.1. Centralized Feedback Scheme

Performing a transform of  $\mathbf{x}(n)$  by using an orthogonal matrix  $\mathbf{T}$ , we obtain the received signal vector at the transform domain as

$$\mathbf{x}_T(n) = \mathbf{T} \mathbf{x}(n) \quad (20)$$

with

$$\mathbf{x}_T(k) = [(\mathbf{x}_T^{(1)}(k))^T (\mathbf{x}_T^{(2)}(k))^T \dots (\mathbf{x}_T^{(M)}(k))^T]^T, \quad (21)$$

where  $\mathbf{x}_T^{(m)}(n)$  is the signal vector at the  $m$ th transform domain bin. Denote  $\mathbf{w}_T = [(\mathbf{w}_T^{(1)})^T (\mathbf{w}_T^{(2)})^T \dots (\mathbf{w}_T^{(M)})^T]^T$  as the weight vector in the transform domain. Then the output of the transform domain array system becomes

$$y_T(k) = \mathbf{w}_T^T \mathbf{x}_T(k) = \mathbf{w}_T^T \mathbf{T} \mathbf{x}(k). \quad (22)$$

Again, using the MMSE criterion

$$\min_{\mathbf{w}_T} E |y_T(k) - d_1(k-v)|^2, \quad (23)$$

the optimum weight vector is given by

$$\mathbf{w}_{T,opt} = \mathbf{R}_T^{-1} \mathbf{r}_T = \mathbf{T}^* \mathbf{w}_{opt}, \quad (24)$$

where

$$\begin{aligned} \mathbf{R}_T &= E[\mathbf{x}_T^*(k)\mathbf{x}_T^T(k)] \\ &= \mathbf{T}^* \mathbf{R} \mathbf{T}^T \\ &= (\mathbf{T} \mathcal{H})^* (\mathcal{H} \mathbf{T})^T + \sigma \mathbf{I}_{MN}, \end{aligned} \quad (25)$$

$$\mathbf{r}_T = E[\mathbf{x}_T^*(k)d_1(n-v)] = \mathbf{T}^* \mathbf{r} = (\mathbf{T} \mathcal{H}_1)^* \mathbf{e}_v. \quad (26)$$

It is easy to verify that the transform domain array with centralized feedback scheme provides the same steady-state MMSE performance, as given by equation (18). The centralized feedback scheme is depicted in Fig. 1.

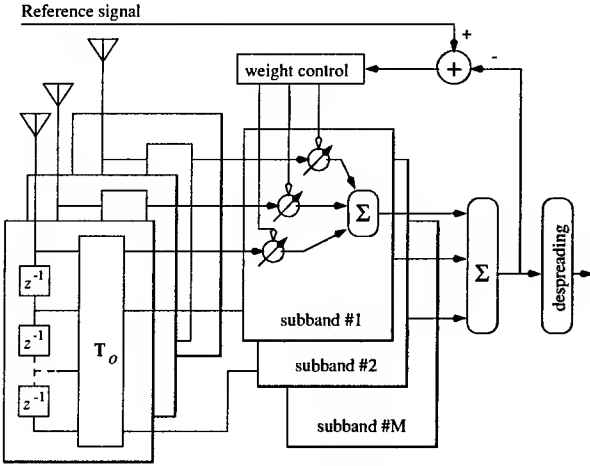


Fig. 1 Subband array with centralized feedback.

### 3.2. Localized Feedback Scheme

We note that the orthogonal transform can reduce the correlation between different transform bins. DFT, filter banks, and wavelets are commonly used methods for providing orthogonal transforms. Here we consider the DFT as the example. Denote

$$\mathbf{T}_o = \begin{bmatrix} W_M^0 & W_M^0 & W_M^0 & \cdots & W_M^0 \\ W_M^0 & W_M^1 & W_M^2 & \cdots & W_M^{M-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ W_M^0 & W_M^{M-1} & W_M^{2(M-1)} & \cdots & W_M^{(M-1)^2} \end{bmatrix} \quad (27)$$

as the  $M \times M$  transform matrix at the output of each array sensor, where

$$W_M = \exp\left(\frac{-2\pi j}{M}\right), \quad (28)$$

then the transform matrix  $\mathbf{T}$  becomes

$$\mathbf{T} = \mathbf{P}_2(\mathbf{I}_N \otimes \mathbf{T}_o)\mathbf{P}_1, \quad (29)$$

where  $\otimes$  denotes the Kronecker product. In (29),  $\mathbf{P}_1$  is a permutation matrix to change the order of the vector  $\mathbf{x}(n)$  such that the  $M$  samples at each array sensor align together, and  $\mathbf{P}_2$  is another permutation matrix that allows the  $N$  data of each bin to align together.

$\mathbf{T}$  can be expressed in the form

$$\mathbf{T} = \begin{bmatrix} \mathbf{T}^{(1)} \\ \vdots \\ \mathbf{T}^{(M)} \end{bmatrix}, \quad (30)$$

where  $\mathbf{T}^{(m)}$  is the  $N \times NM$  submatrix of the matrix  $\mathbf{T}$  corresponding to the  $m$ th bin. Denote

$$\mathbf{x}^{(m)}(n) = \mathbf{T}^{(m)}\mathbf{x}(n) \quad (31)$$

as the signal vector at the  $m$ th subband. When the signal correlation between different transform bins is small,

we ignore the off-block-diagonal elements of the correlation matrix  $\mathbf{R}_T$ , yielding an approximation by the block-diagonal matrix

$$\mathbf{R}'_T = \begin{bmatrix} \mathbf{R}_T^{(1)} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_T^{(2)} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \vdots & \mathbf{R}_T^{(M)} \end{bmatrix} \quad (32)$$

where

$$\begin{aligned} \mathbf{R}_T^{(m)} &= E[(\mathbf{x}_T^{(m)}(n))^* (\mathbf{x}_T^{(m)}(n))^T] \\ &= (\mathbf{T}^{(m)}\mathbf{H})^* (\mathbf{T}^{(m)}\mathbf{H})^T + \sigma^2 \mathbf{I}_N \end{aligned} \quad (33)$$

is the signal covariance matrix of  $\mathbf{x}_T^{(m)}(n)$ . Using the property of block-diagonal matrix, we have

$$(\mathbf{R}'_T)^{-1} = \begin{bmatrix} (\mathbf{R}_T^{(1)})^{-1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & (\mathbf{R}_T^{(2)})^{-1} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \vdots & (\mathbf{R}_T^{(M)})^{-1} \end{bmatrix}. \quad (34)$$

Therefore, the inversion computation of dimension  $NM \times NM$  becomes  $M$  parallel group of matrix inversion of dimension  $N \times N$ , as such the computations can be greatly reduced. When recursive methods are used, it is realized by using  $M$  parallel control loops with  $N$  weights in each loop. The localized feedback scheme is shown in Fig. 2.

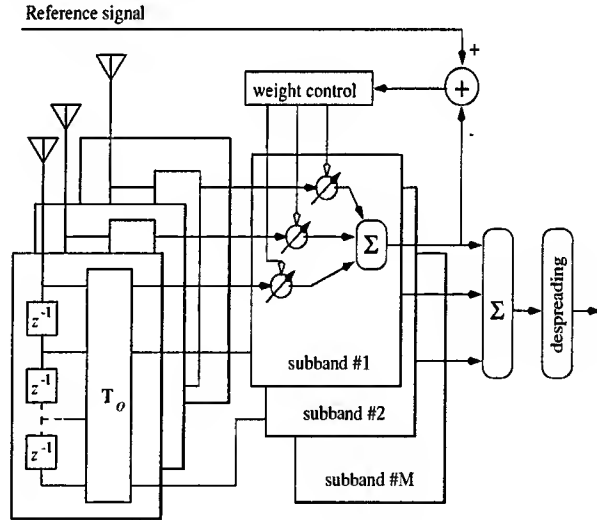


Fig. 2 Subband array with localized feedback.

We use  $d_1(k)$  as the reference signal at each transform bin. In this case, the cross-correlation vector between the received signal vector and the reference signal at the  $m$ th transform bin becomes

$$\mathbf{r}_T^{(m)} = E[(\mathbf{x}_T^{(m)}(k))^* d_1(k-v)] = [\mathbf{T}^{(m)}]^* \mathbf{r}. \quad (35)$$

In the localized feedback scheme, the weight vector at each bin can be obtained from the  $N \times N$  correlation matrix  $\mathbf{R}_T^{(m)}$  and the  $N \times 1$  correlation vector  $\mathbf{r}_T^{(m)}$  which are determined only by the data vector and reference signal at that bin, i.e.,

$$\mathbf{w}_T^{(m)} = (\mathbf{R}_T^{(m)})^{-1} \mathbf{r}_T^{(m)}. \quad (36)$$

Therefore, the centralized feedback transform domain array can be approximated by a set of parallel independent rank-reduced adaptive array processors at each bin, at the cost of ignoring the correlation between signals at different bins. Such transform domain array with the localized feedback scheme can be easily implemented by using a set of parallel array processors, each with the number of weights equal to  $N$ , instead of  $NM$ .

It is clear that

$$\mathbf{r}'_T = \left[ \left( \mathbf{r}_T^{(1)} \right)^T \left( \mathbf{r}_T^{(2)} \right)^T \cdots \left( \mathbf{r}_T^{(M)} \right)^T \right]^T = \mathbf{r}_T. \quad (37)$$

Therefore, the equivalent full-band weight vector of the localized feedback transform domain array becomes

$$\mathbf{w}'_T = (\mathbf{R}'_T)^{-1} \mathbf{r}'_T = (\mathbf{R}'_T)^{-1} \mathbf{r}_T. \quad (38)$$

The corresponding MSE of the localized feedback scheme is given by

$$\text{MSE}_{LF} = 1 + \mathbf{r}_T^H (\mathbf{R}'_T)^{-1} \mathbf{R}_T (\mathbf{R}'_T)^{-1} \mathbf{r}_T - 2\text{Re} \left[ \mathbf{r}_T^H (\mathbf{R}'_T)^{-1} \mathbf{r}_T \right]. \quad (39)$$

Equation (39) implies that the localized feedback transform domain array approach is suboptimal, and, its performance depends on the significance of the cross-correlation between signals at different bins. It is clear from (25) and (39) that the off-block-diagonal element of matrix  $\mathbf{R}_T$ , and subsequently the MSE performance of the localized feedback subband array, depend on both the transform matrix  $\mathbf{T}$  and the channels  $\mathbf{H}_p, p = 1, 2, \dots, P$ .

### 3.3. Partial Feedback Scheme

In the previous subsection, we discussed the transform domain array with localized feedback scheme as an approximation of the transform domain array with centralized feedback scheme. Such localized feedback scheme reduces the number of weights at each bin at the expense of performance reduction, since the off-block-diagonal elements are not considered in the weight estimation.

A subband array with partial feedback, which is shown in Fig. 3, is also possible and provides more flexibility in trading-off the system complexity with the steady-state MSE performance. As shown below, the partial feedback scheme is a generalization of the centralized and localized feedback schemes, which can be considered as two extreme and special cases.

In the transform domain array with partial feedback scheme, the total  $M$  bins are divided into  $K$  groups. The number of bins in  $i$ th group is  $M_i, i = 1, 2, \dots, K$ , with  $M_1 + M_2 + \dots + M_K = M$ . In this paper, we consider the simple case of  $M_1 = M_2 = \dots = M_K = M/K$ .

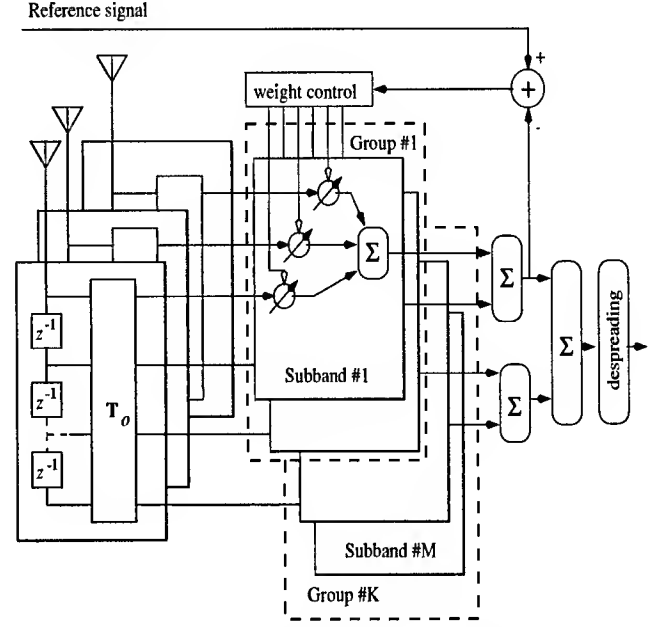


Fig. 3 Subband array with partial feedback.

In this case, the signal covariance matrix  $\mathbf{R}_T$  is approximated by a new block-diagonal matrix  $\mathbf{R}''_T$  with larger block size  $M_1 N$ , expressed as

$$\mathbf{R}''_T = \begin{bmatrix} \mathbf{R}_T^{(G_1)} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_T^{(G_2)} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \vdots & \mathbf{R}_T^{(G_K)} \end{bmatrix} \quad (40)$$

where  $\mathbf{R}_T^{(G_i)}$  is of dimension  $M_1 N \times M_1 N$ . For  $M_1 > 1$ , fewer off-block-diagonal elements are ignored in  $\mathbf{R}''_T$  compared to  $\mathbf{R}'_T$ . Therefore, the partial feedback scheme provides more accurate weights estimation, and subsequently better MSE results, as compared with the localized feedback scheme. Similar to the localized feedback case, the weight vector in the partial feedback scheme is given by

$$\mathbf{w}''_T = (\mathbf{R}''_T)^{-1} \mathbf{r}''_T = \begin{bmatrix} (\mathbf{R}_T^{(G_1)})^{-1} \mathbf{r}_T^{(G_1)} \\ (\mathbf{R}_T^{(G_2)})^{-1} \mathbf{r}_T^{(G_2)} \\ \vdots \\ (\mathbf{R}_T^{(G_K)})^{-1} \mathbf{r}_T^{(G_K)} \end{bmatrix} \quad (41)$$

where

$$\mathbf{r}_T^{(G_i)} = E \left[ \left( \mathbf{x}_T^{(G_i)}(k) \right)^* d_1(k-v) \right], \quad (42)$$

as  $d_1(k-v)$  is used as the reference signal at each group, and

$$\mathbf{x}_T^{(G_i)}(k) = \left[ \left( \mathbf{x}_T^{((i-1)M_1+1)}(k) \right)^T \cdots \left( \mathbf{x}_T^{(iM_1)}(k) \right)^T \right]^T. \quad (43)$$

Since

$$\mathbf{r}_T'' = \left[ \left( \mathbf{r}_T^{(G_1)} \right)^T \quad \left( \mathbf{r}_T^{(G_2)} \right)^T \quad \dots \quad \left( \mathbf{r}_T^{(G_K)} \right)^T \right]^T = \mathbf{r}_T, \quad (44)$$

the MSE of the partial feedback array is therefore

$$\text{MSE}_{PF} = 1 + \mathbf{r}_T^H (\mathbf{R}_T'')^{-1} \mathbf{R}_T (\mathbf{R}_T'')^{-1} \mathbf{r}_T - 2\text{Re} [\mathbf{r}_T^H (\mathbf{R}_T'')^{-1} \mathbf{r}_T]. \quad (45)$$

It is noted that, the partial feedback scheme simplifies to the centralized feedback scheme when  $M_1 = M$ . In this case,  $\mathbf{R}_T''$  becomes  $\mathbf{R}_T$ , and equation (45) becomes equation (18). On the other hand, the localized feedback scheme is achieved by setting  $M_1 = 1$ . In this case,  $\mathbf{R}_T''$  becomes  $\mathbf{R}_T'$ , and equation (45) becomes equation (39).

#### 4. CONVERGENCE PERFORMANCE

In this section, we consider the convergence performance of the transform domain arrays with centralized feedback and localized feedback. The popularly used least mean square (LMS) algorithm is considered.

One of the key factors affecting the convergence performance in the proposed transform domain arrays is the number of controllable weights in the feedback system. In the transform domain array with centralized feedback scheme, the number of weights is  $NM$ , whereas in the cases of the transform domain array with localized feedback and partial feedback schemes, the number of weights in each independent control loop is  $N$  and  $M_1N$ , respectively (although the number of total weights of the entire bins remains  $NM$ ).

It is known that the convergence rate of LMS algorithm depends on the eigenvalue spread, i.e., the ratio between the maximum and minimum eigenvalues of the covariance matrix [8]. Since the covariance matrix defined at a bin,  $\mathbf{R}_T^{(m)}$ ,  $m = 1, \dots, M$ , or that defined at several bins,  $\mathbf{R}_T^{(G_i)}$ ,  $i = 1, \dots, K$ , is a submatrix of  $\mathbf{R}_T$ , from the interlacing property [9], the eigenvalue spread of  $\mathbf{R}_T^{(m)}$  and that of  $\mathbf{R}_T^{(G_i)}$  are smaller than that of  $\mathbf{R}_T$ . Therefore, the transform domain arrays with localized and partial feedback provide improved convergence performance.

On the other hand, when comparing the STAP system and the transform domain array with centralized feedback scheme, since an orthonormal transform does not change the eigenvalues, it is clear that the eigenvalue spread of  $\mathbf{R}$  and  $\mathbf{R}_T$  are the same. Therefore, the STAP system and the centralized feedback transform domain array offer the same convergence performance [6]. However, if the signal powers at different bins are different (due to, e.g., pulse shaping filtering, frequency-selective channel characteristics), the convergence performance can be improved by performing power compensation at the different bins so that the eigenvalue spread is reduced [10, 11, 12].

#### 5. CONCLUSION

We have analyzed the performance of transform domain arrays for DS-CDMA systems with different types of feedback schemes, and derived the respective expressions of the mean square error (MSE). For all proposed schemes,

the transformation is performed in the chip level before despreading. It has been shown that transform domain arrays with localized and partial feedback schemes are generally suboptimal, and their MSE performance depends on the transform matrix of the analysis filters as well as the communication channel characteristics. Since the localized feedback scheme reduces the number of weights at the control loop, the convergence rate is usually improved, which is of practical importance in implementing space-time adaptive processing in fast fading environments. The partial feedback scheme generalizes the other two proposed schemes, namely, the centralized and localized feedback systems. This scheme provides the flexibility to balance the system complexity with the steady-state and convergence performance.

#### REFERENCES

- [1] A. J. Paulraj and C. B. Papadakis, "Space-time processing for wireless communications," *IEEE Signal Processing Magazine*, vol. 14, no. 6, pp. 49-83, Nov. 1997.
- [2] U. Madhow and M. Honig, "MMSE interference suppression for direct-sequence spread-spectrum CDMA," *IEEE Trans. Commun.*, vol. 42, pp. 3178-3188, Dec. 1994.
- [3] H. Liu and M. D. Zoltowski, "Blind equalization in antenna array CDMA systems," *IEEE Trans. Signal Processing*, vol. 45, no. 1, pp. 161-172, Jan. 1997.
- [4] U. Madhow, "Blind adaptive interference suppression for direct-sequence CDMA," *Proc. IEEE*, vol. 86, no. 10, pp. 2049-2069, Oct. 1998.
- [5] Y. Zhang, K. Yang, and M. G. Amin, "Adaptive subband arrays for multipath fading mitigation," in *Proc. IEEE AP-S Int. Symp.*, Atlanta, GA, pp. 380-383, June 1998.
- [6] Y. Kamiya and Y. Karasawa, "Performance comparison and improvement in adaptive arrays based on the time and frequency domain signal processing," *IEICE Trans. Commun.*, vol. J82-A, no. 6, pp. 867-874, June 1999.
- [7] G. Strang and T. Nguyen, *Wavelets and Filter Banks*, Wellesley-Cambridge, 1996.
- [8] S. Haykin, *Adaptive Filter Theory*, 3rd Ed. New Jersey: Prentice Hall, 1996.
- [9] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 3rd Ed. Maryland: John Hopkin Univ. Press, 1996.
- [10] J. C. Lee and C. K. Un, "Performance analysis of frequency-domain block LMS adaptive digital filters," *IEEE Trans. Circuits and Systems*, vol. 36, no. 2, pp. 173-189, Feb. 1989.
- [11] M. de Courville and P. Dujamel, "Adaptive filtering in subbands using weighted criterion," *IEEE Trans. Signal Processing*, vol. 46, no. 9, pp. 2359-2371, Sept. 1998.
- [12] K. Yang, Y. Zhang, and Y. Mizuguchi, "Subband realization of space-time adaptive processing for mobile communications," in *Proc. 10th Int. Symp. on Personal, Indoor and Mobile Radio Communications*, Osaka, Sept. 1999.

# SECTORIZED SPACE-TIME ADAPTIVE PROCESSING FOR CDMA SYSTEMS

*Kehu Yang<sup>1</sup>, Yoshihiko Mizuguchi<sup>1</sup>, and Yimin Zhang<sup>2</sup>*

<sup>1</sup> ATR Adaptive Communications Research Laboratories,  
Seika-cho, Soraku-gun, Kyoto 619-0288, Japan  
Email: yang@acr.atr.co.jp

<sup>2</sup> Department of Electrical and Computer Engineering,  
Villanova University, Villanova, PA 19085, USA  
Email: zhang@ece.vill.edu

## ABSTRACT

Space-time adaptive processing (STAP) is an effective technique of suppressing both the multiuser access interference (MUAI) and the inter-symbol interference (ISI) in wideband CDMA mobile communication systems. However, its complexity is one of the key problems in practical implementations. In this paper we propose adaptive antenna techniques that realize low-complexity space-time adaptive processing within a given spatial sector by spatial-smoothing subarray beamforming sectorization. The proposed technique has the close performance to that of the associated optimum element-space STAP system.

## I. INTRODUCTION

In direct-sequence code-division multiple-access (DS-CDMA) systems, adaptive antennas under the scheme of space-time adaptive processing (STAP) [1, 2] is called as two-dimensional RAKE (2-D RAKE) receivers [3], and is known to be an effective method in suppressing both the multiuser access interference (MUAI) and the inter-symbol interference (ISI). However, the prohibitive computation complexity of STAP systems is one of the key problems in the practical implementations which restricts their application to practical systems and To reduce their complexity, optimal and sub-optimal approaches based on parallel implementation and low-rank transformations have been proposed so far [4-8].

Beamspace-based partially adaptive processing methods are the sub-optimal approaches widely used in array signal processing, where reduced-dimension processing is performed via employing a few beams to encompass the significant components in the systems [4, 9]. The sectorized beamspace adaptive diversity combiner is one of the applications which is effective in combating multipath fading in the wireless communications [4]. References [5] and [6] proposed other two approaches that involve the

wideband beamforming and the reduced-dimension beamforming, respectively.

In this paper we propose novel low-complexity sectorized adaptive antenna techniques which use the spatial-smoothing subarray beamformers to achieve effective beam diversity as well as sufficient degrees of freedom (DOF's) for MUAI suppression. In the proposed techniques, the full field of view is divided into a number of spatial sectors, wherein the sectorized STAP is performed individually. The array is partitioned into a set of subarrays, each forms a beam to cover the same specific sector of interest. In the sector of interest, the number of MUAI's is greatly reduced from the full field-of-view condition. The sectorized STAP scheme combines the advantages of the reduced-rank beamspace processing and the spatio-temporal processing techniques. In comparison with the conventional STAP systems performed in the full field of view, the complexity of the sectorized processing is highly reduced whereas the performance loss to that of the optimum STAP systems can be kept small.

## II. ARRAY SIGNAL MODEL

Consider a cellular CDMA base station using an antenna array of  $N$  ( $N > 1$ ) elements with  $P$  users. The  $p$ -th user's baseband waveform of the transmitted signal is expressed as

$$s_p(t) = \sum_{m=-\infty}^{+\infty} \bar{s}_p(m) \rho_p(t - mT), \quad (1)$$

where  $\bar{s}_p(m)$  denotes the  $m$ -th information symbol of the  $p$ -th user,

$$\rho_p(t) = \sum_{j=0}^{N_c-1} c_p(j) \psi(t - jT_c), 0 \leq t \leq T \quad (2)$$

represents the signature waveform of the  $p$ -th user,  $\{c_p(j)\}_{j=0}^{N_c-1}$  is the spreading code assigned to the  $p$ -th user,  $N_c$  is the number of chips per symbol,  $\psi(t)$  is the



normalized chip waveform limited within  $[0, T_c]$ , and  $T_c$  is the chip interval. The spreading sequence can be periodic or aperiodic, which depends on the standard to be used. In this paper, we consider the periodic case, i.e., the non-random CDMA systems.

The array receiving signal vector  $\mathbf{x}(t)$  is denoted as

$$\begin{aligned} \mathbf{x}(t) &= \sum_{p=1}^P \sum_{l=1}^{L_p} \mathbf{a}(\theta_l^p) \xi_l^p s_p(t - \tau_l^p) + \mathbf{n}(t) \\ &= \sum_{p=1}^P \sum_{m=-\infty}^{+\infty} \bar{s}_p(m) \mathbf{g}_p(t - mT) + \mathbf{n}(t) \end{aligned} \quad (3)$$

where

$$\mathbf{g}_p(t) = \sum_{l=1}^{L_p} \mathbf{a}(\theta_l^p) \xi_l^p \rho_p(t - \tau_l^p), \quad (4)$$

$\{\theta_l^p, \tau_l^p, \xi_l^p\}$  express respectively the angle-of-arrival (AOA), the time delay, and the propagation loss corresponding to the  $l$ -th path of the  $p$ -th user. Moreover,  $\mathbf{a}(\theta)$  is the array steering vector corresponding to  $\theta$ ;  $\bar{s}_p(m)$  denotes  $m$ -th information symbol of the  $p$ -th user,  $L_p$  is the total number of multipath rays of the  $p$ -th user,  $T = N_c T_c$  is the symbol duration, and  $\mathbf{n}(t)$  is the array noise vector.

Define

$$\mathbf{h}_p(t) = \sum_{l=1}^{L_p} \mathbf{a}(\theta_l^p) \xi_l^p \psi(t - \tau_l^p) \quad (5)$$

as the channel response of the  $p$ -th user, we can rewrite (3) as

$$\mathbf{x}(t) = \sum_{p=1}^P \sum_{m=-\infty}^{+\infty} \sum_{j=0}^{N_c-1} \bar{s}_p(m) c_p(j) \mathbf{h}_p(t - jT_c - mT) + \mathbf{n}(t). \quad (6)$$

We make the following assumptions:

A1) The information symbols  $\bar{s}_p(m)$ ,  $p = 1, \dots, P$  are i.i.d., and satisfy  $E\{\bar{s}_p(m) \bar{s}_q^*(n)\} = \delta_{p,q} \delta_{m,n}$ , where  $(\cdot)^*$  denotes complex conjugation and  $\delta_{p,q}$  denotes the Kronecker delta function.

A2) The channels  $\{\mathbf{h}_p(t), p = 1, \dots, P\}$  are linear and time-invariant with a finite duration within  $[0, D_p T_c]$ . Here, we assume  $D_p T_c > T$  for wideband CDMA channels.

A3) The noise vector is zero-mean, temporally and spatially white with  $E\{\mathbf{n}(t) \mathbf{n}^T(t)\} = 0$  and  $E\{\mathbf{n}(t) \mathbf{n}^H(t)\} = \sigma^2 \mathbf{I}$ , where  $(\cdot)^T$  and  $(\cdot)^H$  denote transpose and conjugate transpose, respectively,  $\sigma^2$  expresses the noise power, and  $\mathbf{I}$  is the identity matrix. The noise vector is also assumed to be uncorrelated with the user

signals.

Denote  $\Delta = T_c/J$  as the sampling cycle, where  $J \geq 1$  is an integer which expresses the factor of oversampling. Thus, sampling at  $t = i\Delta + nT_c$ , the discrete form of (5) becomes

$$\begin{aligned} \mathbf{x}(i\Delta + nT_c) &= \sum_{p=1}^P \sum_{m=-\infty}^{+\infty} \sum_{j=0}^{N_c-1} \bar{s}_p(m) c_p(j) \times \\ &\quad \mathbf{h}_p(i\Delta + nT_c - jT_c - mT) + \mathbf{n}(i\Delta + nT_c) \\ &\quad i=0, \dots, J-1. \end{aligned} \quad (7)$$

By stacking  $\mathbf{x}(i\Delta + nT_c)$ ,  $i=0, \dots, J-1$ , we have

$$\underline{\mathbf{x}}(n) = \sum_{p=1}^P \sum_{d=0}^{D_p} \mu_p(n-d) \underline{\mathbf{h}}_p(d) + \underline{\mathbf{n}}(n), \quad (8)$$

where  $\mu_p(n)$  is the chip-rate signal sequence of the  $p$ -th user. In (8), we use the notation  $\underline{\alpha}(n) = [\alpha^T(nT_c), \dots, \alpha^T(nT_c + (J-1)\Delta)]^T$ , where  $\alpha$  denotes either  $\mathbf{x}$ ,  $\mathbf{h}$  or  $\mathbf{n}$ .

### III. SYMBOL-LEVEL PROCESSING

#### 1. Chip-level optimum adaptive processing

For the consecutive samples during the period of  $M$  chips ( $M > N_c$ ), we form the following vectors

$$\mathbf{X}(n) = [\underline{\mathbf{x}}^T(n), \underline{\mathbf{x}}^T(n-1), \dots, \underline{\mathbf{x}}^T(n-M+1)]^T, \quad (9)$$

$$\mathbf{S}_p(n) = [\mu_p(n), \mu_p(n-1), \dots, \mu_p(n-M-D_p+1)]^T, \quad (10)$$

$$\mathbf{N}(n) = [\underline{\mathbf{n}}^T(n), \underline{\mathbf{n}}^T(n-1), \dots, \underline{\mathbf{n}}^T(n-M+1)]^T. \quad (11)$$

Define the following Sylvester convolution matrix of user  $p$  by the impulse response of its vector channel,  $[\underline{\mathbf{h}}_p^T(0), \underline{\mathbf{h}}_p^T(1), \dots, \underline{\mathbf{h}}_p^T(D_p)]^T$ , as

$$\mathbf{H}_p^{(M)} = \begin{bmatrix} \underline{\mathbf{h}}_p(0) & \dots & \underline{\mathbf{h}}_p(D_p) & \mathbf{0} & \dots & \dots & \mathbf{0} \\ \mathbf{0} & \underline{\mathbf{h}}_p(0) & \dots & \underline{\mathbf{h}}_p(D_p) & \mathbf{0} & \dots & \mathbf{0} \\ \vdots & \ddots & \dots & \ddots & \dots & \ddots & \vdots \\ \mathbf{0} & \dots & \dots & \mathbf{0} & \underline{\mathbf{h}}_p(0) & \dots & \underline{\mathbf{h}}_p(D_p) \end{bmatrix} \quad (12)$$

with the dimension of  $MN_c \times (M+D_p)$ , and (8) is extended to

$$\mathbf{X}(n) = \sum_{p=1}^P \mathbf{H}_p^{(M)} \mathbf{S}_p(n) + \mathbf{N}(n). \quad (13)$$

The output of the STAP under (13) is described as,

$$\mathbf{y}(n) = \mathbf{W}^T \mathbf{X}(n). \quad (14)$$

Under the minimum mean square error (MMSE) criterion

$$\min_w E \left| \mu_{p_0}(n-v) - y(n) \right|^2, \quad (15)$$

where  $\mu_{p_0}(n)$  is the training chip sequence of the user  $p_0$ , which is considered as the desired user, and  $v \geq 0$  is the delay of the training signal selected to minimize the MMSE. The optimum weights are given by the Wiener-Hopf equation as

$$W_{p_0, \text{chip}}^* = \mathbf{R}_X^{-1} \mathbf{r}_{p_0}(v), \quad (16)$$

where

$$\mathbf{R}_X = E[X(n)X^H(n)] \quad (17)$$

is the space-time correlation matrix, and

$$\mathbf{r}_{p_0}(v) = E[\mu_{p_0}^*(n-v)X(n)] \quad (18)$$

expresses the cross-correlation vector between the training signal and the received signal vector. It is seen that the complexity of the chip-level adaptive filter depends on the dimension of the signal vector, i.e., the dimension of the weight vector that is selected based on the length of the associated channels.

It is noted that in CDMA systems, the performance of the chip-level processing is confined to the number of the degrees of freedom (DOF's) provided by the employed array and the cyclostationarity of the users' signals. Such a problem can be mitigated in the scheme of symbol-level processing, where the MUAI components become quasi-random noises after despreading with the signature code of the desired user.

## 2. Symbol-level optimum adaptive processing

Symbol-level processing is so called that symbol-duration spaced taps are used in the space-time filter. Similar to the oversampling-based subchannel formulation as made in (7), and (8), the subchannel-based signal vector after despreading the array receiving signals with the signature code of the desired user  $p_0$  is denoted as

$$X_c(mN_c) \triangleq \sum_{l=0}^{N_c-1} X_s(mN_c + l)c_{p_0}(l), \quad (19)$$

where

$$X_s(\beta) = [\mathbf{x}^T(\beta), \mathbf{x}^T(\beta-1), \dots, \mathbf{x}^T(\beta-N_c+1)]^T. \quad (20)$$

By stacking  $K$  consecutive-symbol samples, we have the space-time signal vector as

$$\bar{X}_c(m) = [X_c^T(mN_c), \dots, X_c^T((m-K+1)N_c)]^T \quad (21)$$

Let  $M=KN_c$ , from (19)-(21), it is seen that  $\bar{X}_c(m)$  has the same form as (13). This implies

$$\begin{aligned} \bar{X}_c(m) = & \sum_{p=1}^P \mathbf{H}_p^{(M)} \sum_{l=0}^{N_c-1} S_p(mN_c + l)c_{p_0}(l) \\ & + \sum_{l=0}^{N_c-1} N(mN_c + l)c_{p_0}(l). \end{aligned} \quad (22)$$

It is seen that  $\sum_{l=0}^{N_c-1} S_p(mN_c + l)c_{p_0}(l)$  has  $KN_c + D_{p_0}$

components that are the consecutive samples of the single-path despreading signal waveform plotted in Fig. 1, where the peaks are the desired finger outputs. The peak

components of the vector  $\sum_{l=0}^{N_c-1} S_p(mN_c + l)c_{p_0}(l)$ ,  $p \neq p_0$

standing for the MUAI's should be suppressed because they could lead to false fingers in the situations where the near-far problem exists. When there is no near-far problem, they are considered as quasi-random noises. The symbol-level adaptive processing can be performed based on (21), i.e.,

$$y_c(m) = W^T \bar{X}_c(m) = \sum_{l=0}^{K-1} \mathbf{w}_l^T X_c((m-l)N_c), \quad (23)$$

where  $W = [\mathbf{w}_0^T, \mathbf{w}_1^T, \dots, \mathbf{w}_{K-1}^T]^T$ . Similar to the chip-level processing, under the symbol-level MMSE criterion

$$\min_w E \left| \bar{s}_{p_0}(m-v) - y_c(m) \right|^2, \quad (24)$$

the optimum weight vector is obtained as

$$W_{p, \text{symbol}}^* = \mathbf{R}_c^{-1} \underline{\gamma}_{p_0}(v), \quad (25)$$

where

$$\mathbf{R}_c = E[\bar{X}_c(m)\bar{X}_c^H(m)], \quad (26)$$

$$\underline{\gamma}_{p_0}(v) = E[\bar{s}_{p_0}^*(m-v)\bar{X}_c(m)], \quad (27)$$

$\bar{s}_{p_0}(m)$  denotes the training symbol sequence of  $p_0$ -th user, and  $v$  is selected in the same way as explained in (15).

It is noted that the above filter (25) still has the same complexity as that given in (16).

## IV. SECTORIZED SPACE-TIME ADAPTIVE PROCESSING

### 1. Lower-rank beamspace transformation

Lower-rank beamspace transformation is known to be an effective way to reduce the complexity of an array processing system. Unlike the scheme of the conventional beamforming, here we consider the smoothing subarray beamforming illustrated in Fig. 2.

Define  $\mathbf{b} = [b_1, b_2, \dots, b_{N-\kappa}]^T$  as the beamformer vector, which forms a beam to encompass the desired signal at each of the  $\kappa-1$  subarrays ( $\kappa < N$ ). Then, the output signal vector

of the beamforming in Fig. 2 is denoted as

$$\mathbf{x}_b(t) = \mathbf{B}^T \mathbf{x}(t) \quad (28)$$

where  $\mathbf{x}_b(t) = [x_{b1}(t), x_{b2}(t), \dots, x_{bk}(t)]^T$ , and the beamformer matrix  $\mathbf{B}$  is expressed by

$$\mathbf{B} = \begin{bmatrix} b_1 & 0 & \dots & 0 \\ \vdots & b_1 & \ddots & \vdots \\ b_{N-K} & \vdots & \ddots & 0 \\ 0 & b_{N-K} & \ddots & b_1 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & b_{N-K} \end{bmatrix}_{N \times (K+1)} \quad (29)$$

## 2 Sectorized space-time adaptive processing

The sectorized space-time adaptive processing can be performed in the same way as that described in Section II and III by replacing  $\mathbf{x}(t)$  with  $\mathbf{x}_b(t)$ . Define

$$\mathbf{z}_p(t) = \mathbf{B}^T \mathbf{h}_p(t) \quad (30)$$

and

$$\mathbf{n}_b(t) = \mathbf{B}^T \mathbf{n}(t) \quad (31)$$

$$\mathbf{x}_b(i\Delta + nT_c) = \mathbf{B}^T \mathbf{x}(i\Delta + nT_c), \quad i=0, \dots, J-1, \quad (32)$$

we have

$$\underline{\mathbf{x}}_b(n) = \sum_{p=1}^P \sum_{d=0}^{D_p} \mu_p(n-d) \underline{\mathbf{z}}_p(d) + \underline{\mathbf{n}}_b(n), \quad (33)$$

where

$$\underline{\mathbf{x}}_b(n) = [\mathbf{x}_b^T(nT_c), \dots, \mathbf{x}_b^T(nT_c + (J-1)\Delta)]^T, \quad (34)$$

$$\underline{\mathbf{z}}_p(n) = [\mathbf{z}_p^T(nT_c), \dots, \mathbf{z}_p^T(nT_c + (J-1)\Delta)]^T, \quad (35)$$

$$\underline{\mathbf{n}}_b(n) = [\mathbf{n}_b^T(nT_c), \dots, \mathbf{n}_b^T(nT_c + (J-1)\Delta)]^T. \quad (36)$$

By stacking the  $N_c$  consecutive samples, we have

$$\mathbf{X}_b(n) = [\underline{\mathbf{x}}_b^T(n), \underline{\mathbf{x}}_b^T(n-1), \dots, \underline{\mathbf{x}}_b^T(n-N_c+1)]^T. \quad (37)$$

The symbol-level vector after despreading the output signals vector  $\mathbf{X}_b(n)$  can be denoted as

$$\mathbf{X}_{bc}(mN_c) \triangleq \sum_{l=0}^{N_c-1} \mathbf{X}_b(mN_c + l) \mathbf{c}_{p_0}(l). \quad (38)$$

Similar to (23), the symbol-level sectorized space-time adaptive processing can be performed as

$$y_{bc}(mN_c) = \mathbf{W}^T \bar{\mathbf{X}}_{bc}(mN_c) = \sum_{l=0}^{K-1} \mathbf{w}_l^T \mathbf{X}_{bc}((m-l)N_c), \quad (39)$$

where

$$\bar{\mathbf{X}}_{bc}(m) = [\mathbf{X}_{bc}^T(mN_c), \dots, \mathbf{X}_{bc}^T((m-K+1)N_c)]^T. \quad (40)$$

Under MMSE criterion

$$\min_{\mathbf{W}} E |\bar{s}_{p_0}(m-v) - y_{bc}(mN_c)|^2, \quad (41)$$

the optimum weight vector is obtained as

$$\mathbf{W}_{p_0, \text{sector}}^* = \mathbf{R}_{bc}^{-1} \gamma_{p_0}^{(b)}(v), \quad (42)$$

where

$$\mathbf{R}_{bc} = E[\bar{\mathbf{X}}_{bc}(mN_c) \bar{\mathbf{X}}_{bc}^H(mN_c)], \quad (43)$$

$$\gamma_{p_0}^{(b)}(v) = E[\bar{s}_{p_0}^*(m-v) \bar{\mathbf{X}}_{bc}(mN_c)], \quad (44)$$

$\bar{s}_{p_0}(m)$  and  $v$  are of the same meaning as that in (27), respectively.

To further reduce the complexity, we can use only the significant components over a threshold within the vector  $\mathbf{X}_{bc}(mN_c)$ , as is commonly implemented. We denote it as the simplified scheme. The results of the simplified scheme are included and compared at the computer simulations.

## V. COMPUTER SIMULATIONS

Computer simulations are performed to confirm the effectiveness of the proposed techniques. In these simulations, an eight-element uniform linear array with half-wavelength spacing is used. The array is partitioned into subarrays, and beams are formed at each subarray. For example, the beamformer for a three-subarray partitioning (six array sensors at each subarray) can be designed as

$$\mathbf{b} = [e^{-j1.25u}, e^{-j0.75u}, e^{-j0.25u}, e^{j0.25u}, e^{j0.75u}, e^{j1.25u}]^T$$

where  $u = 2\pi \sin(\theta^0)$  and  $\theta^0$  dictates the central angle of the sector where the spatial rays of the desired user signals are located. In the simulations, 18 CDMA users' signals are considered to be present, where user 1 is considered as the desired user. The code length of all the users is 127. Each user has 6 multipath rays. It is assumed that the AOA's of the paths are Gaussian distributed for each user, and their propagation loss and time delay obey the Rayleigh and the exponential distributions, respectively. Detailed parameters for the desired user are given in Table 1. The signal-to-noise ratio (SNR) of the direct ray of the user 1 is assumed as -10dB, and the SNR's of the direct rays of the other users are randomly chosen from -12.7 dB to -6.6 dB. And their nominal AOA's are uniformly distributed. The central angle of the given sector is assumed as  $\theta^0 = 12.3^\circ$ .

We selected  $K=2$ , i.e., two taps for the symbol-level space-time adaptive processing. The steady state residual error powers of the normal sectorized STAP and its simplified scheme are plotted in Fig. 3, respectively, where

the number of subarrays is changed from one to four. In the simplified scheme, the threshold is taken as 1.8 times the standard deviation of the components' amplitudes of the signal vector  $X_{bc}(mN_c)$ . The residual error power of the element-space STAP is  $-25.36\text{dB}$ , which is considered as the bound of the sectorized processing and is also plotted in Fig. 3. It is clear that the results of three-beam and four-beam sector STAP are close to the bound, whereas the complexity and the computational burden are greatly reduced, especially for the simplified scheme with the acceptable performance loss.

## VI. CONCLUSIONS

We have proposed sectorized STAP techniques for CDMA systems, which provide effective sub-optimal low-complexity implementation of a STAP system. Simulation results show close performance to the optimal element-space STAP system.

## REFERENCES

- [1] A. J. Paulraj and C. B. Papadias, "Space-time processing for wireless communications," *IEEE Signal Processing Magazine*, vol. 14, no. 6, pp. 49-83, Nov. 1997.
- [2] R. Kohno, "Spatial and temporal communication theory using adaptive antenna array," *IEEE Personal Communications*, vol. 5, no. 1, pp. 28-35, Feb. 1998.
- [3] H. Liu and M. D. Zoltowski, "Blind equalization in antenna array CDMA systems," *IEEE Trans. Signal Processing*, vol. 45, no. 1, pp. 161-172, Jan. 1997.
- [4] T.-S. Lee and Z. S. Lee, "A sectorized beamspace adaptive diversity combiner for multipath environments", *IEEE Trans. Vehi. Technol.*, vol. 48, pp. 1503-1510, Sept. 1999.
- [5] J. Ramos, M. D. Zoltowski, and H. Liu, "Low-complexity space-time processing for DS-CDMA communications", *IEEE Trans. Signal Processing*, vol. 48, no. 1, Jan. 2000.
- [6] Y.-F. Chen, M. D. Zoltowski, J. Ramos, C. Chatterjee, and V. P. Roychowdhury, "Reduced-dimension blind space-time 2-D Rake receivers for DS-CDMA communication systems," *IEEE Trans. Signal Processing*, vol. 48, no. 6, June. 2000.
- [7] K. Yang, Y. Zhang, and Y. Mizuguchi, "Spatio-temporal signal subspace-based subband space-time adaptive processing," in *Proc. Int. Symp. on Antennas and Propagation*, Fukuoka, Japan, Aug. 2000.
- [8] Y. Zhang, K. Yang, and M. G. Amin, "Transform domain array processing for CDMA systems," in *Proc. IEEE Workshop on Statistical Signal and Array Processing*, Pocono Manor, PA, Aug. 2000.
- [9] B. D. Van Veen and R. A. Roberts, "Partially adaptive beamformer design via output power minimization," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-35, pp. 1524-1532, 1987.

Table 1 Parameters of the desired user

No.	$\theta$ (deg.)	$\pi(\text{sym.})$	$\xi$
1	12.3	0	$0.045 + 0.998i$
2	7	0.02	$0.93 - 0.206i$
3	11.1	0.32	$-0.803 - 0.160i$
4	14.2	1.33	$-0.459 + 0.424i$
5	11.6	1.81	$0.355 - 0.264i$
6	26.2	1.82	$-0.264 + 0.034i$

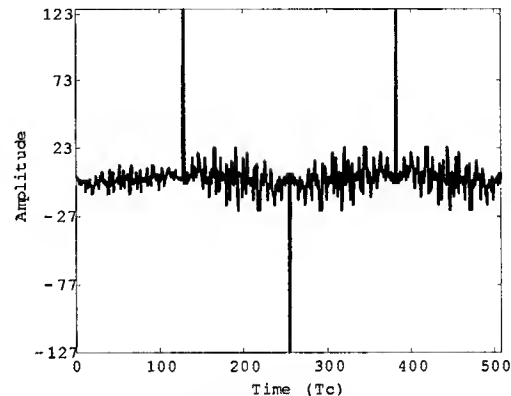


Fig. 1 Single-path despreading waveform

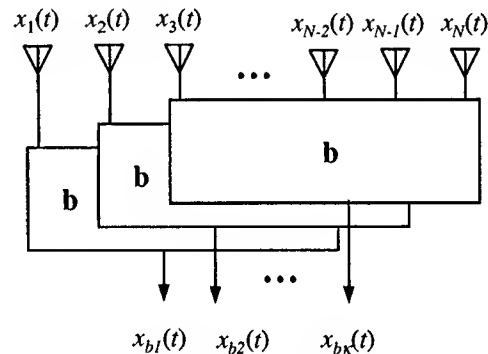


Fig. 2 Smoothing subarray beamforming

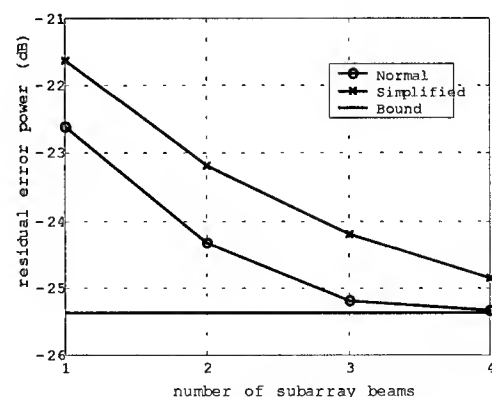


Fig. 3 Residual error power

# DEMODULATION OF AMPLITUDE MODULATED SIGNALS IN THE PRESENCE OF MULTIPATH

*Zhengyuan Xu and Ping Liu*

Dept. of Electrical Engineering  
University of California  
Riverside, CA 92521  
{dxu, pliu}@ee.ucr.edu

## ABSTRACT

Signals modulated by  $M$ -ary pulse amplitude modulation (PAM) or  $M$ -ary quadrature amplitude modulation (QAM) have certain structured constellation. When the communication channel introduces inter-symbol interference (ISI) at the receiver end, demodulation of such signals can be performed by constant modulus algorithm (CMA) based equalizers to cancel the interference. However, characteristics of modulated signals are only partially considered in the CMA cost function. In this paper, more constraints are imposed on the equalized signal to fully capture the property of the modulated signal both in its phase and amplitude. Observing that PAM signals are uniformly spaced on the x-axis and QAM signals in two-dimensional signal space, the property of transmitted signals from each category can be included in an equivalent deterministic mathematical description, similar to the constant modulus. This description is absorbed in our modified cost function, resulting in a simultaneous minimization of dispersion relevant to signal's phase and amplitude. The performance of the equalizers based on these new algorithms are compared with the CMA equalizer.

## 1. INTRODUCTION

In different wireless applications, different modulation schemes are employed to meet specific resource or service requirements. Each modulation exhibits its own property. Signals by  $M$ -ary pulse amplitude modulation (PAM) or  $M$ -ary quadrature amplitude modulation (QAM) have certain structured constellation. For PAM signals, they are uniformly spaced in the real axis (x-axis), while QAM signals are uniformly distributed in a 2-dimensional signal space. If such signals are transmitted through a multipath channel, signal demodulation requires an equalizer to mitigate the channel distortion. The particular source characteristics of-

ten facilitate the equalizer design. The constant modulus algorithm (CMA) based equalizer is widely used [7] and shows its unique capability in equalizing signals with constant modulus property [5]. It was first proposed by [3]. Extensive studies on such equalizers have followed [1], [2], [4]. The algorithm minimizes the deviation of modulus of equalized signal from a constant. The satisfactory performance can be achieved especially when the transmitted signal has constant modulus property.

It seems that the knowledge about the phase of the modulated signal is dismissed in CMA. However, this knowledge plays an equivalent role in many cases in representing a signal. It can be expected that its incorporation into the cost function will improve the equalization performance. To equalize a dispersive channel (could be complex) with  $M$ -PAM transmitted signals, the dispersion in the distance of the equalized signal away from the x-axis should also be minimized together with its modulus deviation. Similarly, when a  $M$ -QAM signals are transmitted, it is not sufficient to consider only the amplitude of the equalized signal in a 2-dimensional signal space, since they are uniformly distributed along both directions which are perpendicular to each other and parallel to two axes. Motivated by CMA algorithm, we will design new equalizers for these two kinds of modulated signals by taking into account their equally spaced property in our new cost function. Similar to CMA algorithm, the stochastic gradient descent methods are employed to update our equalizers. The performance of the equalizers based on these new algorithms are compared with the CMA equalizer.

## 2. PROBLEM STATEMENT

In wireless communications, the multipath channel introduces inter-symbol interference (ISI) in the received

signal  $\mathbf{x} \in C^n$  [4]

$$\mathbf{x} = \mathbf{H}\mathbf{s} + \mathbf{w} \quad (1)$$

where  $\mathbf{s} \in C^m$  is the complex source vector from either  $M$ -PAM or  $M$ -QAM constellation,  $\mathbf{H} \in C^{p \times m}$  is a complex channel matrix,  $\mathbf{w} \in C^p$  represents additive white Gaussian noise (AWGN), and  $\mathbf{x} \in C^p$  is the received signal vector. To detect the signal  $\mathbf{s}(l)$ , an equalizer  $\mathbf{f} \in C^p$  is designed. Its output  $y$  can be written as

$$y = \mathbf{f}^H \mathbf{x} = \mathbf{a}^T \mathbf{s} + \mathbf{f}^H \mathbf{w} \quad (2)$$

where superscripts  $(\cdot)^T$ ,  $(\cdot)^H$  stand for transpose and Hermitian respectively,  $\mathbf{a}^T = \mathbf{f}^H \mathbf{H}$  is the composite response of the channel and the equalizer. Perfect equalization can be achieved in the absence of noise if the equalizer can compensate the channel in such a way that  $\mathbf{a}$  has only one non-zero element [6]

$$\mathbf{a} = e^{j\theta} [0, \dots, 0, 1, 0, \dots, 0]^T \quad (3)$$

Therefore the output will be a delayed input with some phase shift. ISI is completely eliminated in the absence of noise. Different criteria can be used to seek perfect equalization. In CMA criterion, the dispersion of the modulus of equalizer output about a constant is minimized

$$J_c(\mathbf{f}) \triangleq E\{(|y|^2 - r_0)^2\} \quad (4)$$

where “ $E$ ” represents expectation,  $r_0 = \frac{E\{|s(l)|^4\}}{E\{|s(l)|^2\}}$ . The algorithm is usually implemented by stochastic gradient descent method

$$\mathbf{f}(k+1) = \mathbf{f}(k) - \mu(|y(k)|^2 - r_0)y^*(k)\mathbf{x}(k) \quad (5)$$

where  $*$  represents conjugate. It is clear that the modulus characteristic is captured and employed. However, most modulated signals possesses properties in both amplitudes and phase. The  $M$ -PAM or  $M$ -QAM signals take discrete values from a set whose elements lie on the x-axis or a 2-dimensional signal space uniformly. Motivated by CMA criterion, we will derive a new cost function to incorporate this information and develop a corresponding algorithm to obtain the equalizer next.

### 3. PROPOSED EQUALIZERS

Let us first review the representations and properties of PAM and QAM signals. The PAM signals are one dimensional in the sense that they are real and uniformly distributed on the real axis. The QAM signals are complex and uniformly spaced in directions of real axis and imaginary axis. Due to this similarity, the

properties of QAM can be easily found once the properties of PAM signals are explored. For a general discussion, the multipath channel and the equalizer are assumed to be complex for both cases. We start with the equalization of PAM signals.

#### 3.1. PAM signals

$M$ -ary PAM signals can be represented by the following

$$\tilde{s}_m = (2\tilde{m} - 1 - M)d, \quad \tilde{m} = 1, \dots, M$$

where  $\tilde{m}$  is a random number. Usually  $M$  is an even integer and can be written as  $M \triangleq 2L$ . These PAM signals can also be expressed by

$$s_m = (2m - 1)d, \quad m = -L, \dots, L \quad (6)$$

if we define a new variable  $m \triangleq \tilde{m} - L$ . We will adopt this signal description later. In (6),  $m$  can only take integers from  $-L$  to  $L$  which can be expressed by  $s_m$  as:  $m = \frac{s_m + d}{2d}$ . In the current context, this constraint is equivalent to  $\sin(m\pi) = 0$ . Thus it requires

$$\sin\left(\frac{s_m + d}{2d}\pi\right) = \cos\left(\frac{s_m}{2d}\pi\right) = 0 \quad (7)$$

The transformation from (6) to (7) is essential in constructing our cost function. The other property of  $s_m$  is that it has phase equal to a multiple of  $\pi$  because  $s_m$  lies on the real axis. Therefore

$$\sin\phi = 0 \quad (8)$$

where  $\phi$  is the phase of  $s_m$ . Taking into account the complex equalized signal, we can combine (4), (7) and (8) in one cost function

$$J_1(\mathbf{f}) = E\{(|y|^2 - r_0)^2 + \alpha_1 \cos^2\left(\frac{|y|}{2d}\pi\right) + \alpha_2 \sin^2(\hat{\phi})\} \quad (9)$$

where  $\alpha_1$  and  $\alpha_2$  are weighting factors,  $y$  is the equalized signal given by (2),  $\hat{\phi}$  is its phase. In (9),  $y$  and  $\hat{\phi}$  are functions of our equalizer  $\mathbf{f}$ . Therefore  $J_1(\mathbf{f})$  is a highly non-linear function of  $\mathbf{f}$  and difficult to minimize. Similar to CMA algorithm, we update the equalizer according to gradient descent method

$$\mathbf{f}(k+1) = \mathbf{f}(k) - \mu_1 \nabla J_1(\mathbf{f})|_{\mathbf{f}=\mathbf{f}(k)} \quad (10)$$

The derivative of  $J_1(\mathbf{f})$  with respect to  $\mathbf{f}^H$  is required in (10). It can be derived term by term from the RHS of (9). The first term is directly from CMA. Its derivative can be easily found to be

$$(E\{(|y|^2 - r_0)^2\})'_f = 2E\{(|y|^2 - r_0)y^* \mathbf{x}\} \quad (11)$$

For the second term, its derivative can be computed once derivatives of  $|y|$  and  $\hat{\phi}$  are obtained. If we express  $|y|$  by  $\sqrt{yy^*}$ , then the derivative of  $|y|$  is easily computed to be

$$(|y|)'_f = \frac{y^*}{2|y|}x \quad (12)$$

For  $\hat{\phi}$ , it can be expressed by  $\mathbf{f}$  as

$$\hat{\phi} = \arctan \frac{y - y^*}{j(y + y^*)} = \arctan \frac{\mathbf{f}^H \mathbf{x} - \mathbf{x}^H \mathbf{f}}{j(\mathbf{f}^H \mathbf{x} + \mathbf{x}^H \mathbf{f})} \quad (13)$$

Therefore the derivative of  $\hat{\phi}$  can be shown to be

$$(\hat{\phi})'_f = \frac{\mathbf{x}^H \mathbf{f}}{2j|y|^2} \mathbf{x} = \frac{1}{2jy} \mathbf{x} \quad (14)$$

According to (9), (11), (12) and (14), the derivative of  $\nabla J_1(\mathbf{f})$  is obtained as

$$\nabla J_1(\mathbf{f}) = E\{\beta \mathbf{x}\} \quad (15)$$

where

$$\beta = 2(|y|^2 - r_0)y^* + \alpha_2 \frac{\sin(2\hat{\phi})}{2jy} - \alpha_1 \frac{\pi y^* \sin(\frac{|y|\pi}{d})}{4d|y|}$$

Therefore the stochastic gradient algorithm for the equalizer follows

$$\mathbf{f}(k+1) = \mathbf{f}(k) - \mu_1 \beta \mathbf{x} \quad (16)$$

### 3.2. QAM signals

There are some similarities between PAM and QAM signals. In the signal space QAM signals can be depicted by  $(s_x, s_y)$  where

$$s_x = (2m_x - 1)d_x, \quad m_x = -L_x, \dots, L_x \quad (17)$$

$$s_y = (2m_y - 1)d_y, \quad m_y = -L_y, \dots, L_y \quad (18)$$

This representation can be transformed into (see (7))

$$\cos(\frac{s_x}{2d_x}\pi) = 0, \quad \cos(\frac{s_y}{2d_y}\pi) = 0 \quad (19)$$

Therefore we can build the following cost function

$$J_2(\mathbf{f}) = E\{\cos^2(\frac{y_1}{2d_x}\pi) + \cos^2(\frac{y_2}{2d_y}\pi)\} \quad (20)$$

with  $y_1$  and  $y_2$  to be real and imaginary parts of  $y$  respectively. The gradient descent recursion for the equalizer can be formulated as

$$\mathbf{f}(k+1) = \mathbf{f}(k) - \mu_2 \nabla J_2(\mathbf{f})|_{\mathbf{f}=\mathbf{f}(k)} \quad (21)$$

To compute  $\nabla J_2(\mathbf{f})$ , we first evaluate derivatives of  $y_1$  and  $y_2$ . If they are expressed explicitly by  $\mathbf{f}$ ,

$$y_1 = \frac{y + y^*}{2} = \frac{\mathbf{f}^H \mathbf{x} + \mathbf{x}^H \mathbf{f}}{2}$$

$$y_2 = \frac{y - y^*}{2j} = \frac{\mathbf{f}^H \mathbf{x} - \mathbf{x}^H \mathbf{f}}{2j}$$

then it is easy to show that their derivatives have the form

$$(y_1)'_f = \frac{\mathbf{x}}{2}, \quad (y_2)'_f = \frac{\mathbf{x}}{2j}$$

Based on these results and (20),  $\nabla J_2(\mathbf{f})$  can be derived to be

$$\nabla J_2(\mathbf{f}) = -E\{\eta \mathbf{x}\} \quad (22)$$

where

$$\eta = \frac{\pi \sin(\frac{y_1}{d_x}\pi)}{2d_x} + \frac{\pi \sin(\frac{y_2}{d_y}\pi)}{2jd_y}$$

In the case  $d_x = d_y = 1$ ,  $\eta$  is simplified to

$$\eta = \frac{\pi}{2}[\sin(y_1\pi) - j\sin(y_2\pi)]$$

Substituting (22) in (21) and using instantaneous approximation, we can update the equalizer according to

$$\mathbf{f}(k+1) = \mathbf{f}(k) + \mu_2 \eta \mathbf{x} \quad (23)$$

The equalization method proposed for either PAM source or QAM source in this section explicitly considers the phase and modulus properties of the transmitted signals. As a result, superior performance is expected compared with the conventional CMA equalizer which only captures the modulus property.

## 4. SIMULATIONS

In this section we provide some simulation examples to demonstrate the applicability of the proposed PAM and QAM equalization methods. We also compare them with the CMA algorithm [3] respectively based on inter-symbol interference(ISI) and the error probability. The ISI is used to illustrate the convergence property of the algorithm and defined as

$$ISI = \frac{\sum_l |\mathbf{a}_l|^2 - |\mathbf{a}|_{max}^2}{|\mathbf{a}|_{max}^2}$$

where  $\mathbf{a}^T = \mathbf{f}^H \mathbf{H}$ ,  $|\mathbf{a}|_{max}$  is the largest absolute value of all elements in  $\mathbf{a}$ . Under perfect equalization,  $\mathbf{a}$  has only one nonzero component as in (4). Then ISI becomes zero. Therefore, small ISI indicates the proximity to the desired response. To gain more insight about the performance of the methods in the communications context, we also adopt error probability as

the other measure. It is defined as the percentage of accumulated decoding errors among total number of transmitted symbols up to the current iteration, and obtained from multiple independent realizations with random input signals.

In the experiments, we consider an unknown non-minimum phase channel impulse response used in [6] with the first 4 coefficients  $[-0.400 \ 0.840 \ 0.336 \ 0.134]$ . The equalizer has 12 taps and the initial value of all zeros  $[0, \dots, 0, 1, 0, \dots, 0]^T$  except that the seventh element is 1. 5000 iterations are run in each realization. Totally 50 independent realizations are performed to obtain the average results.

First, we compare the proposed PAM equalizer with Godard approach [3] with the PAM source. The input signals take six equi-probable values:  $\{\pm 0.1, \pm 0.3, \pm 0.5\}$ . The step size  $\mu$  is set to be 0.085, weighting factors  $\alpha_1 = 0.005$  and  $\alpha_2 = 0$  (since a real channel is used). The first 500 iterations are used for initialization for both methods. The average ISI after 500 iterations is plotted in Fig. 1. The solid line represents the proposed PAM method while the dashed line for CMA. It is observed that the ISI of the proposed PAM method converges to a level 15dB lower than that from CMA while maintaining the same fast convergence. The error probability is also shown by Fig. 2. In fact, based on our observation, the proposed method doesn't take any error after convergence (800 iterations), while CMA still accumulates some errors.

Our second experiment considers QAM source with 4 equi-probable values  $\{\pm 1 \pm j\}$ . We also compares the proposed QAM scheme with the CMA algorithm [3]. The first 20 data points are used for initialization for both methods. The average ISI and error probability after 20 iterations are plotted in Fig. 3 and Fig. 4 respectively. Solid lines represent the proposed QAM equalization method while dashed lines for CMA. It is seen that the ISI based on the proposed QAM scheme converges faster than that of the standard CMA while achieving a much lower level after convergence. The error probability of the proposed method is also much lower than that of CMA. This fact can be reflected by the difference in constellation diagrams of the equalized outputs for all iterations from a randomly-picked realization, as shown in Fig. 5 and Fig. 6. It is interesting to note that the equalized outputs of our equalizer has a much smaller variation than that of the CMA equalizer.

## 5. REFERENCES

- [1] Z. Ding, C.R. Johnson and R.A. Kennedy, "On the (non)existence of undesirable equilibria of Godard equalizers", *IEEE Trans. on Signal Processing*, vol. 40, pp. 2425-2432, Oct. 1992.
- [2] G.J. Foschini, "Equalization without altering or detecting data", *AT&T Tech. J.*, vol. 64, no. 8, pp. 1885-1911, Oct. 1985.
- [3] D.N. Godard, "Self-recovering equalization and carrier tracking in two dimensional data communication systems", *IEEE Trans. on Comm.*, vol. 28, no. 11, pp. 1167-75, November 1980.
- [4] H. Zeng, L. Tong and C.R. Johnson, "An Analysis of Constant Modulus Receivers", *IEEE Trans. on Signal Processing*, vol. 47, no. 11, pp. 2990-1999, November 1999.
- [5] C.R. Johnson, *et.al*, "Blind Equalization Using Constant Modulus Criterion: A Review", *Proc. of the IEEE*, vol. 86, no. 10, pp. 1927-1950, October 1998.
- [6] O. Shalvi and E. Weinstein, "New criteria for blind deconvolution of nonminimum phase systems (channels)", *IEEE Transactions on Information Theory*, vol.36, no.2, pp.312-21, March 1990.
- [7] J.R. Treichler, I. Fijalkow and C.R. Johnson, "Fractionally spaced equalizers", *IEEE Signal Processing Mag.*, pp. 45-81, May 1996.



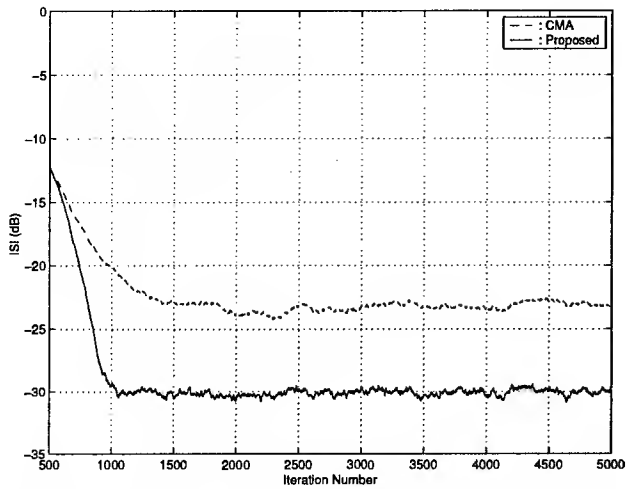


Figure 1: ISI of the proposed method and Godard's method with PAM sources.

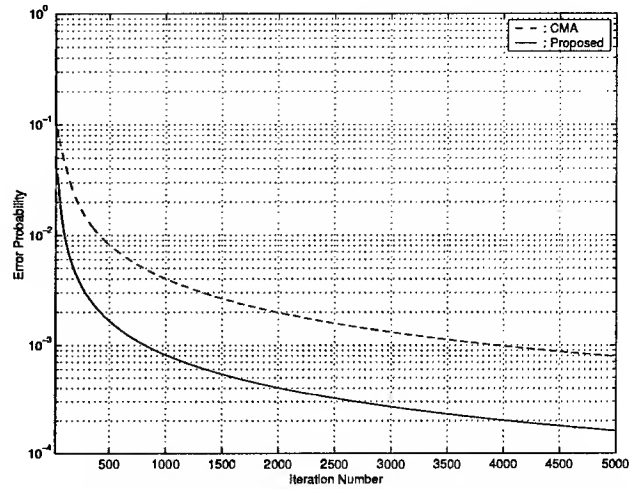


Figure 4: Error probability of the proposed method and Godard's method with QAM sources.

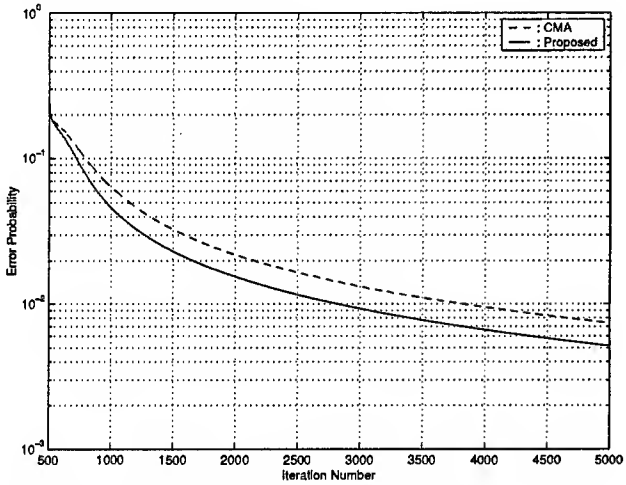


Figure 2: Error probability of the proposed method and Godard's method with PAM sources.

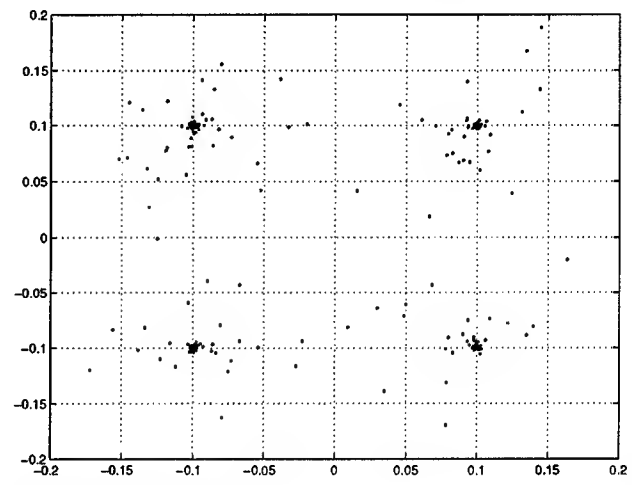


Figure 5: Equalized output of the proposed method with QAM sources.

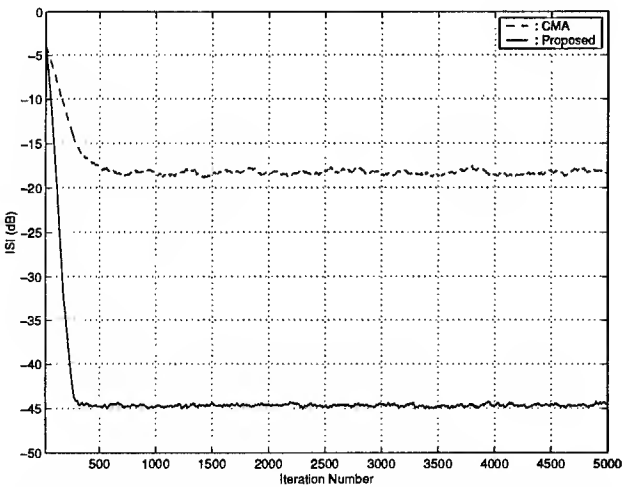


Figure 3: ISI of the proposed method and Godard's method with QAM sources.

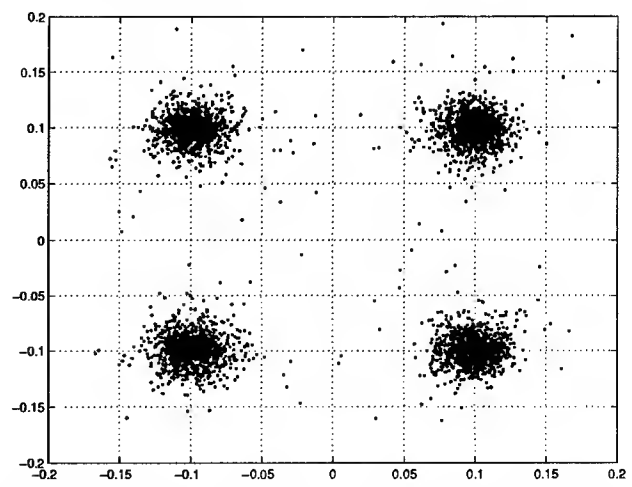


Figure 6: Equalized output of Godard's method with QAM sources.

# MULTICHANNEL AND BLOCK BASED PRECODING METHODS FOR FIXED POINT EQUALIZATION OF NONLINEAR COMMUNICATION CHANNELS

Arthur J. Redfern

Texas Instruments  
12500 TI Boulevard, MS 8653  
Dallas, TX 75243  
redfern@ti.com

G. Tong Zhou

Georgia Institute of Technology  
School of ECE  
Atlanta, GA 30332-0250  
gtz@ece.gatech.edu

## ABSTRACT

Substantial power efficiency improvements are possible in communication systems if a moderate amount of nonlinearity is permitted at the transmitter amplifier and corrected for at the receiver. The Volterra series is a suitable model for many power amplifiers, and is readily incorporated into communication channel models. Existing fixed point equalization algorithms for Volterra channels place restrictive conditions on the locations of first-order kernel zeros. We show that multichannel and block based precoding linear equalization techniques can be combined with the fixed point equalizer to allow for exact equalization of Volterra systems with mixed-phase first-order kernels.

## 1. INTRODUCTION

The design of a communication system, from the data format to the transceivers, is composed of many parts. Radio frequency power amplifier design is an important component of cellular, television, radio, and data transmission systems. In amplifier design the requirements of power efficiency and linearity can be at odds with each other, with the result being that power efficiency is sacrificed in order to meet linearity requirements [2].

Substantial efficiency improvements can be possible if some mild nonlinearity is allowed in the transmitter amplifier and corrected for at the receiver. This improved efficiency translates to lower operating costs, longer battery life, and smaller size devices. A penalty of allowing additional nonlinearity into the system is that the equalizer must now compensate for a nonlinear channel.

In this paper we consider fixed point equalization of communication channels modeled by the Volterra series [3], [4], [8]. Fixed point equalization in this case

refers to the contraction mapping theorem [3] (not integer arithmetic). The Volterra series is a useful nonlinear model for amplifiers [2], and is readily incorporated into the overall channel model as an extension of linear convolution.

Drawbacks of traditional fixed point equalization techniques include the requirement that the linear component of the channel is minimum-phase (for stable exact inverses) [3] or its zeros are not near the unit circle (for approximate inverses) [4]. These can be serious limitations for realistic communication channel models, as the error in the inversion of the linear channel component is iterated on by the fixed point algorithm.

Recently, multichannel [7] and block based precoding methods [5] have become popular for linear channel equalization. This is because both methods convert the ill-posed single channel inversion problem into a well posed problem with an exact (zero forcing) solution in the noise-free case. We show that these principles can be combined with the fixed point equalizer, for zero forcing equalization of nonlinear channels with mixed-phase first-order kernels.

## 2. THE VOLTERRA SERIES

For the discrete  $J$ th-order Volterra system  $H$ , the input  $x(n)$  is related to the output  $y(n)$  by:

$$\begin{aligned} y(n) &= H(x(n), \dots, x(n-L)) \\ &= \sum_{j=1}^J H_j(x(n), \dots, x(n-L_j)) \\ &= \sum_{j=1}^J \sum_{\tau_1=0}^{L_j} \cdots \sum_{\tau_j=\tau_{j-1}}^{L_j} h_j(\tau_1, \dots, \tau_j) \prod_{a=1}^j x(n-\tau_a), \end{aligned}$$

where  $H_j$  is the  $j$ th-order operator of  $H$ ,  $h_j(\tau_1, \dots, \tau_j)$  is the nonredundant region of the  $j$ th-order kernel, and

This work was supported in part by NASA grant NGT-352334 and NSF grant MIP-9703312.

$L = \max\{L_1, \dots, L_J\}$ . Notice that a first-order Volterra system ( $J = 1$ ) is linear convolution (an FIR filter).

Throughout this paper the symbol  $u(n)$  will be used to refer to the linear component of the Volterra system with additive noise  $v(n)$ :

$$u(n) = \sum_{\tau_1=0}^{L_1} h_1(\tau_1)x(n - \tau_1) + v(n).$$

For the channel input  $x(n)$ , output  $y(n)$ , noise  $v(n)$ , and linear portion of the output with noise  $u(n)$ , it will be assumed that these vectors are composed of a basic block of  $N$  symbols, and a subscript will indicate how many symbols before this basic block to include, e.g.:

$$\mathbf{x}_L = [x(-L), \dots, x(N-1)]^T.$$

An optional argument can be included to specify a subset of the vector:

$$\mathbf{x}_L(a : b) = [x(a), \dots, x(b)]^T.$$

If the  $d$  sample delay operator  $z^{-d}$  is placed before the vector, then each element of the vector is delayed by  $d$  samples:

$$z^{-d}\mathbf{x}_L = [x(-L-d), \dots, x(N-1-d)]^T.$$

We define the Volterra series relationship between an input vector  $\mathbf{x}_L$  and output vector  $\mathbf{y}_0$  as:

$$\mathbf{y}_0 = H(\mathbf{x}_L).$$

As a shorthand notation to refer to the output of specific order operators, we define:

$$H_{a:b}(\mathbf{x}_L) = \sum_{j=a}^b H_j(\mathbf{x}_{L_j}).$$

It is often necessary to write the first-order operator corresponding to a finite impulse response (FIR) filter as a filtering matrix. For the length  $Q+1$  vector  $\mathbf{c} = [c(Q), \dots, c(0)]^T$ , the  $N \times N+Q$  filtering matrix  $\mathcal{T}_N(\mathbf{c})$  is defined as:

$$\mathcal{T}_N(\mathbf{c}) = \begin{bmatrix} c(Q) & \cdots & c(0) & & \\ & \ddots & & \ddots & \\ & & c(Q) & \cdots & c(0) \end{bmatrix}.$$

### 3. FIXED POINT EQUALIZATION

In this section we review the single channel fixed point equalizer based on the contraction mapping theorem. The basic idea underlying fixed point equalization of

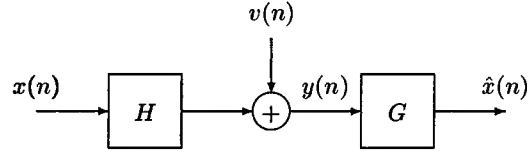


Figure 1: A single channel Volterra system.

Volterra channels is setting up a fixed point equation for the input in terms of the known system kernels and system output, then solving for the input using the method of successive approximations [3]. Two assumptions are implicit:

*Assumption (A1):* The  $K+L$  previous input symbols  $x(-K-L), \dots, x(-1)$  have already been estimated.

*Assumption (A2):* The  $K$  previous output samples  $y(-K), \dots, y(-1)$  are available.

In the following derivation, even though  $\mathbf{x}_0$  will be on the left hand side of the equation and  $\mathbf{x}_{K+L}$  will be on the right hand side, there is still a fixed point equation in  $\mathbf{x}_0$  since  $\mathbf{x}_{K+L}$  can be formed directly from  $\mathbf{x}_0$ .

The derivation of the fixed point equalizer is well known in the literature [3]. Here the derivation is performed using the notation of Section 2 which will emphasize the importance of the inversion of the linear component of the noisy channel output.

The input/output relationship for the single channel Volterra system in Fig. 1 with additive noise at the receiver is:

$$\mathbf{y}_0 = \mathbf{H}_N \mathbf{x}_{L_1} + H_{2:J}(\mathbf{x}_L) + \mathbf{v}_0, \quad (1)$$

where  $\mathbf{H}_n = \mathcal{T}_n(\mathbf{h}_1)$ . Rearranging the terms and applying the linear operator  $G_s$  with memory  $K$  to both sides yields:

$$G_s(\mathbf{H}_{K+N} \mathbf{x}_{K+L_1} + \mathbf{v}_K) = G_s(\mathbf{y}_K) - G_s H_{2:J}(\mathbf{x}_{K+L}). \quad (2)$$

Notice that each of the vectors and matrices from (1) to (2) has been extended by  $K$  samples in the past (available from (A1) and (A2)) since the operator  $G_s$  has memory  $K$ . To setup the desired fixed point equation, it is necessary to make the left hand side of (2)  $z^{-d}\mathbf{x}_0$ . Define the single channel error term as

$$\epsilon_s = z^{-d}\mathbf{x}_0 - G_s(\mathbf{H}_{K+N} \mathbf{x}_{K+L_1} + \mathbf{v}_K)$$

It is common to choose  $G_s$  corresponding to a causal  $K$ th order FIR filter

$$\mathbf{g}_s = [g_s(K), \dots, g_s(0)]^T,$$

designed according to the minimum mean-square error (MMSE) criterion. For the MMSE equalizer it is necessary to make the following assumption:

*Assumption (A3):* The input  $x(n)$  and the noise  $v(n)$  are mutually uncorrelated, stationary random processes with known covariance matrices:

$$\begin{aligned}\mathbf{R}_{xx} &= E[\mathbf{x}_{K+L_1}(n-K-L_1:n) \\ &\quad \mathbf{x}_{K+L_1}^H(n-K-L_1:n)] \\ \mathbf{R}_{vv} &= E[\mathbf{v}_K(n-K:n) \mathbf{v}_K^H(n-K:n)],\end{aligned}$$

respectively.

If (A1) - (A3) are satisfied, then the equalizer  $\mathbf{g}_s$  can be solved for as:

$$\mathbf{g}_s = \mathbf{R}_{uu}^{-T} \mathbf{r}_{xu}, \quad (3)$$

where  $\mathbf{R}_{uu}$  and  $\mathbf{r}_{xu}$  are defined as

$$\begin{aligned}\mathbf{R}_{uu} &= \mathbf{H}_{K+1} \mathbf{R}_{xx} \mathbf{H}_{K+1}^H + \mathbf{R}_{vv} \\ \mathbf{r}_{xu} &= \mathbf{H}_{K+1}^* E[x(n-d) \mathbf{x}_{K+L_1}^*(n-K-L_1:n)].\end{aligned}$$

Substitution of the operator  $G_s$  associated with the filter  $\mathbf{g}_s$  into (2) results in the fixed point equation:

$$z^{-d} \mathbf{x}_0 = G_s(\mathbf{y}_K) - G_s H_{2,J}(\mathbf{x}_{K+L}) + \boldsymbol{\varepsilon}_s.$$

Assuming that  $\boldsymbol{\varepsilon}_s$  is small, it is ignored and the approximate fixed point equation is solved:

$$z^{-d} \mathbf{x}_0 = G_s(\mathbf{y}_K) - G_s H_{2,J}(\mathbf{x}_{K+L}).$$

For the case of  $d = 0$ ,  $\mathbf{x}_{K+L}$  can be determined from  $z^{-d} \mathbf{x}_0$  and (A1). However, when  $d > 0$ , it is not possible to determine the last  $d$  elements of  $\mathbf{x}_{K+L}$ , namely  $x(N-d), \dots, x(N-1)$ . To obtain proper estimates of these last  $d$  symbols in  $z^{-d} \mathbf{x}_0$ , they could be the first symbols estimated in the next block of data.

A drawback of the fixed point equalizer is the error introduced into the fixed point equation associated with the inverse of the first-order kernel. The error depends on the length  $K$ , delay  $d$ , and the location of the zeros of  $H_1$ . The fixed point equalizers in the following two sections eliminate this source of error, and allow for zero forcing equalization of the linear component (along with the nonlinear component) of the channel in the noise-free case.

#### 4. MULTICHANNEL FIXED POINT EQUALIZER

The availability of multiple observations per symbol period at the receiver has become more common in many communication systems. Using a superscript  $(s)$  to denote the channel, the following assumption is made:

*Assumption (A4):* There are no common zeros across all of the linear components  $\{H_1^{(s)}\}_{s=1}^S$  of the channels.

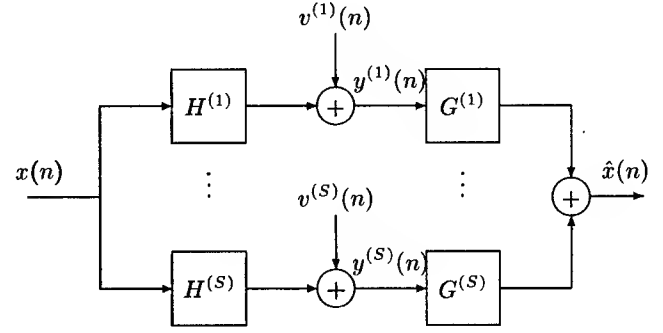


Figure 2: A single-input/multiple-output Volterra system.

It is well known that for multiple linear channels, FIR zero forcing equalization is possible if (A4) is satisfied [7]. In this section it is shown that these linear multichannel equalization techniques can be combined with the fixed point equalizer, to allow for zero forcing equalization of Volterra channels with mixed-phase first-order kernels using as little as two channels.

Consider the multichannel Volterra system shown in Fig. 2 and again assume that (A1) and (A2) are satisfied. For the  $s$ th channel write:

$$\mathbf{y}_0^{(s)} = \mathbf{H}_N^{(s)} \mathbf{x}_{L_1} + H_{2,J}^{(s)}(\mathbf{x}_L) + \mathbf{v}_0^{(s)},$$

where  $\mathbf{H}_n^{(s)} = \mathcal{T}_n(\mathbf{h}_1^{(s)})$ . Rearranging terms and applying the linear operator  $G_m^{(s)}$  with memory  $K$  to both sides yields:

$$\begin{aligned}G_m^{(s)}(\mathbf{H}_{K+N}^{(s)} \mathbf{x}_{K+L_1} + \mathbf{v}_K^{(s)}) \\ = G_m^{(s)}(\mathbf{y}_K^{(s)}) - G_m^{(s)} H_{2,J}^{(s)}(\mathbf{x}_{K+L}).\end{aligned} \quad (4)$$

Because (4) holds for each channel  $s$ , it is possible to sum the results for all  $S$  channels and write:

$$\begin{aligned}\sum_{s=1}^S G_m^{(s)}(\mathbf{H}_{K+N}^{(s)} \mathbf{x}_{K+L_1} + \mathbf{v}_K^{(s)}) = \\ \sum_{s=1}^S G_m^{(s)}(\mathbf{y}_K^{(s)}) - \sum_{s=1}^S G_m^{(s)} H_{2,J}^{(s)}(\mathbf{x}_{K+L}).\end{aligned} \quad (5)$$

If it is possible to make the left hand side of (5)  $\mathbf{x}_0$ , then the result will be the desired fixed point equation. Define the error term as

$$\boldsymbol{\varepsilon}_m = \mathbf{x}_0 - \sum_{s=1}^S G_m^{(s)}(\mathbf{H}_{K+N}^{(s)} \mathbf{x}_{K+L_1} + \mathbf{v}_K^{(s)}).$$

If (A4) is satisfied, then in the noise-free case a  $K$ th-order FIR zero forcing solution exists such that

$$\sum_{s=1}^S G_m^{(s)}(\mathbf{H}_{K+N}^{(s)} \mathbf{x}_{K+L_1}) = \mathbf{x}_0,$$

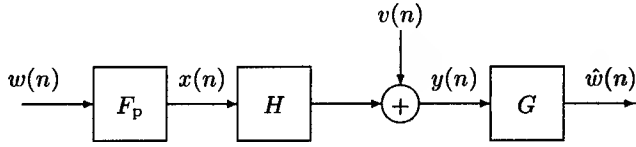


Figure 3: A single channel Volterra system with precoder.

provided that  $S(K+1) \geq K+L+1$  [7]. Define the multichannel filtering matrix  $\mathbf{H}_{m,K+1}$  and the multichannel  $K$ th-order equalizer  $\mathbf{g}_m$  corresponding to  $\{G_m^{(s)}\}_{s=1}^S$  as:

$$\begin{aligned}\mathbf{H}_{m,K+1} &= [\mathbf{H}_{K+1}^{(1),T}, \dots, \mathbf{H}_{K+1}^{(S),T}]^T \\ \mathbf{g}_m &= [\mathbf{g}_m^{(1),T}, \dots, \mathbf{g}_m^{(S),T}]^T.\end{aligned}$$

The zero forcing equalizer can be recovered as [7]:

$$\mathbf{g}_m = (\mathbf{H}_{m,K+1}^T)^{\dagger} \mathbf{e}_{K+L+1}, \quad (6)$$

where  $\mathbf{e}_{K+L+1}$  is a  $(K+L+1) \times 1$  vector with a one in the  $(K+L+1)$ th position and zeros elsewhere.

Substitution of the operator  $G_m^{(s)}$  associated with the filter  $\mathbf{g}_m^{(s)}$  designed according to (6) into (5) results in the fixed point equation:

$$\mathbf{x}_0 = \sum_{s=1}^S G_m^{(s)}(\mathbf{y}_K^{(s)}) - \sum_{s=1}^S G_m^{(s)} H_{2:J}^{(s)}(\mathbf{x}_{K+L}) + \varepsilon_m.$$

## 5. BLOCK BASED PRECODING FIXED POINT EQUALIZATION

As an alternative to using multiple channels at the receiver to improve the single channel inversion problem, structured redundancy could be introduced at the transmitter. By block precoding at the transmitter and block equalization at the receiver, FIR zero forcing equalization of single channel systems is possible irrespective of the location of channel zeros [5]. As in the multichannel case, these properties can be extended to fixed point equalization of Volterra channels.

Consider the block-based transmission scheme of Fig. 3. At the transmitter, data symbols  $w(n)$  are collected into a block of length  $M$ :

$$\mathbf{w} = [w(0), \dots, w(M-1)]^T,$$

and mapped by the precoder  $F_p$  to the length  $N$  block of channel inputs  $\mathbf{x}_0$ . If the precoder is linear, then it can be represented by the  $N \times M$  matrix  $\mathbf{F}_p$ . The precoder structure is chosen to satisfy the following two assumptions [5]:

*Assumption (A5):* The lengths  $L$ ,  $M$ , and  $N$  satisfy  $N = L + M$ .

*Assumption (A6):*  $\text{rank}(\mathbf{F}_p) = M$ , and the last  $L$  rows of  $\mathbf{F}_p$  are zero.

As a result of (A6),  $\mathbf{F}_p$  can be decomposed as

$$\mathbf{F}_p = \begin{bmatrix} \tilde{\mathbf{F}}_p \\ \mathbf{0}_{L \times M} \end{bmatrix}$$

where the  $M \times M$  matrix  $\tilde{\mathbf{F}}_p$  is nonsingular. Using (A6) it is possible to write:

$$\mathbf{x}_L = \begin{bmatrix} \mathbf{0}_{L \times 1} \\ \tilde{\mathbf{F}}_p \mathbf{w} \\ \mathbf{0}_{L \times 1} \end{bmatrix}.$$

The  $N$  row filtering matrix for the first-order kernel  $\mathbf{H}_n = \mathcal{T}_n(\mathbf{h}_1)$  can be decomposed as

$$\mathbf{H}_N = [\check{\mathbf{H}}_N \tilde{\mathbf{H}}_N \bar{\mathbf{H}}_N],$$

where  $\check{\mathbf{H}}_N$  is  $N \times L$ ,  $\tilde{\mathbf{H}}_N$  is  $N \times M$ , and  $\bar{\mathbf{H}}_N$  is  $N \times L$ . Using these definitions, the input/output relationship for the block-based system with precoding can be written as

$$\mathbf{y}_0 = \tilde{\mathbf{H}}_N \tilde{\mathbf{F}}_p \mathbf{w} + H_{2:J} \left( \begin{bmatrix} \mathbf{0}_{L \times 1} \\ \tilde{\mathbf{F}}_p \mathbf{w} \\ \mathbf{0}_{L \times 1} \end{bmatrix} \right) + \mathbf{v}_0.$$

Rearranging terms and applying the linear operator  $G_p$  to both sides yields:

$$G_p(\tilde{\mathbf{H}}_N \tilde{\mathbf{F}}_p \mathbf{w} + \mathbf{v}_0) = G_p(\mathbf{y}_0) - G_p H_{2:J} \left( \begin{bmatrix} \mathbf{0}_{L \times 1} \\ \tilde{\mathbf{F}}_p \mathbf{w} \\ \mathbf{0}_{L \times 1} \end{bmatrix} \right). \quad (7)$$

If the left hand side of (7) was  $\mathbf{w}$ , then the desired fixed point equation would result. Define the error term as

$$\varepsilon_p = \mathbf{w} - G_p(\tilde{\mathbf{H}}_N \tilde{\mathbf{F}}_p \mathbf{w} + \mathbf{v}_0). \quad (8)$$

If (A5) and (A6) are satisfied, then in the noise-free case, a zero forcing solution  $G_p$  (with matrix form  $\mathbf{G}_p$ ) to (8) exists such that [5]:

$$\mathbf{G}_p \tilde{\mathbf{H}}_N \tilde{\mathbf{F}}_p \mathbf{w} = \mathbf{w}.$$

The zero forcing equalizer can be recovered as [5]:

$$\mathbf{G}_p = \tilde{\mathbf{F}}_p^{-1} \tilde{\mathbf{H}}_N^{\dagger}. \quad (9)$$

Substitution of the operator  $G_p$  associated with the matrix  $\mathbf{G}_p$  designed according to (9) into (7) results in the fixed point equation:

$$\mathbf{w} = G_p(\mathbf{y}_0) - G_p H_{2:J} \left( \begin{bmatrix} \mathbf{0}_{L \times 1} \\ \tilde{\mathbf{F}}_p \mathbf{w} \\ \mathbf{0}_{L \times 1} \end{bmatrix} \right) + \varepsilon_p.$$

## 6. SIMULATIONS

We considered a third-order baseband Volterra system with  $L_1 = 5$  and  $L_3 = 2$ , whose complex kernel coefficient's real and imaginary parts were chosen randomly from  $[-0.5, 0.5]$ , with the third-order kernel scaled by 0.03 such that the nonlinear to linear power ratio is -23 dB. A 16-QAM input was used, and additive white Gaussian noise was present at the channel output. For each data point we generated 100 blocks of  $N = 100$  symbols for 100 different channels.

For the multichannel fixed point simulations we used  $S = 4$  channels and the linear component of the equalizer designed according to (6) with order  $K = 8$ . The single channel fixed point simulations (with and without precoding) used the first of the multichannel fixed point simulations' channels. The standard single channel fixed point equalizer's linear component was designed according to (3) with  $K = 32$  and  $d = 16$ . The linear component of the single channel fixed point equalizer with precoding was designed according to (9), with a data block length of  $M = N - L = 95$  and precoder  $\tilde{\mathbf{F}} = \mathbf{I}_{M \times M}$ . For each of the fixed point equalizers, 5 iterations of their respective fixed point equation were performed.

For our performance metric, we calculate the signal to interference ratio (SIR), defined in terms of the MSE of the equalizer output:

$$\text{SIR} = -10 \log_{10} \text{MSE (dB)},$$

vs. SNR. The SIR allows us to assess the ability of the equalizers to cope with both the noise and the nonlinearity. Fig. 4 compares the output of each of the fixed point equalizers, along with the corresponding outputs of the linear components of the equalizers.

## 7. CONCLUSIONS

In this paper we showed that multichannel and block based precoding linear channel equalization techniques can be combined with the fixed point method for zero forcing equalization of Volterra channels with mixed-phase first-order kernels. Since the fixed point equalizer takes the form of a nonlinear correction added to a linear inverse, it is a practical addition to existing linear channel equalization schemes.

## REFERENCES

- [1] G. Giannakis and E. Serpedin, "Linear multichannel blind equalizers of nonlinear FIR Volterra channels," *IEEE Transactions on Signal Processing*, vol. 45, no. 1, pp. 67-81, 1997.

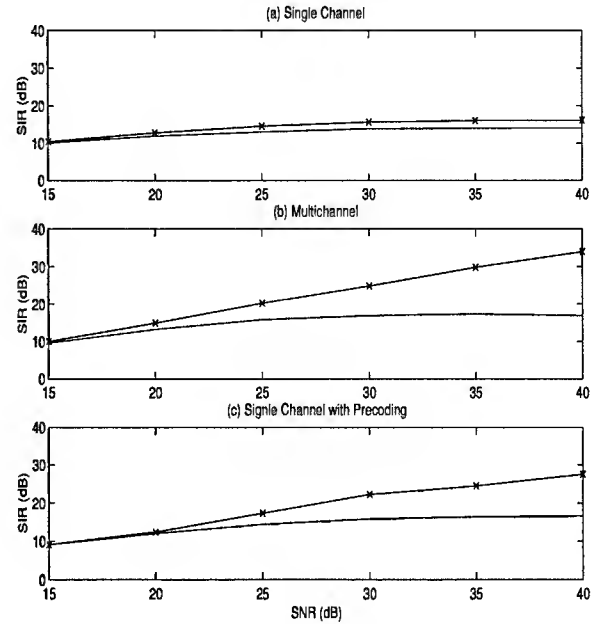


Figure 4: Comparing the linear and fixed point equalizer outputs.

- [2] S. Maas, "Analysis and optimization of nonlinear microwave circuits by Volterra-series analysis," *Microwave Journal*, vol. 33, no. 4, pp. 245-251, Apr. 1990.
- [3] R. Nowak and B. Van Veen, "Volterra filter equalization: A fixed point approach," *IEEE Transactions on Signal Processing*, vol. 45, no. 2, pp. 377-388, 1997.
- [4] A. Redfern and G. Zhou, "A fixed point equalizer for nonlinear communication channels," *Proceedings of the Thirty-Third CISS*, Baltimore, MD, Mar. 1999, to appear.
- [5] A. Scaglione, G. Giannakis and S. Barbarossa "Redundant filterbank precoders and equalizers part I: Unification and optimal designs," *IEEE Transactions on Signal Processing*, vol. 47, no. 7, pp. 1988-2006, 1999.
- [6] M. Schetzen, *The Volterra and Wiener Theories of Nonlinear Systems*. New York: John Wiley and Sons, 1980.
- [7] D. Slock, "Blind fractionally-spaced equalization, perfect-reconstruction filter banks and multichannel linear prediction," *Proceedings of the IEEE ICASSP*, pp. 585-588, Adelaide, Australia, Apr. 1994.
- [8] C. Tseng and E. Powers, "Nonlinear channel equalization in digital satellite systems," *Proceedings of the IEEE Globecom*, pp. 1639-1643, Houston, TX, Nov. 1993.

# JOINT ESTIMATION OF PROPAGATION PARAMETERS IN MULTICARRIER SYSTEMS

Saïd Aouada and Adel Belouchrani

Electrical Engineering Department,  
Ecole Nationale Polytechnique  
P.O. Box 182 El Harrach 16200, Algiers, Algeria  
E-mail: belouchrani@hotmail.com

## ABSTRACT

A joint propagation parameter estimation method for MultiCarrier systems is proposed. The main difference between Single Carrier and MultiCarrier models is outlined and handled in the derivation of the algorithm. The method uses a subspace-based 2-D ESPRIT-like approach, exploiting frequency shift invariance of the system as well as the ULA geometry to provide closed-form estimation. Basic performances of the algorithm are illustrated through simulations and compared with respect to the Cramer-Rao bound.

## 1. INTRODUCTION

In several wireless systems, the transmitted signals are subject to the effects of multipath channels, caused by the remote terrestrial objects and the inhomogeneities in the physical medium. Estimation of the multipath propagation parameters from measurements at a multisensor antenna, provides a better channel characterization for subsequent processing. These parameters include, among others, the Direction Of Arrival (DOA) and Time Difference Of Arrival (TDOA) of each path. In MultiCarrier Modulation (MCM) systems such as Digital Terrestrial Television Broadcasting (DTTB) and Digital Audio Broadcasting (DAB), the transmitted signals are subject to the effects of a multipath channel, in the same way as are Single Carrier Modulation (SCM) systems.

Herein, we investigate the possibility of performing closed-form Joint Angle and Delay Estimation (JADE) for a MCM system in a single batch, in a way similar to JADE for SCM systems, by exploiting the frequency diversity of the system, together with a known array geometry. The system consists of a single source and a single antenna array. A channel model is derived to outline the frequency shift invariance associated with the system. The model exploits the stationarity of the parameters over the coherence time of the channel. It also takes into account the fact that the unknown complex fadings differ from one carrier to another. Both the uniform carrier spacing and a known array geometry allow closed-form estimation of the propagation parameters. More particularly, if the antenna is Uniform Linear (ULA), or has an ESPRIT doublet structure, JADE can be achieved using a 2D ESPRIT-like technique. The Cramer-Rao Bound on the variance of the estimated parameters is also derived from the obtained model.

## 2. DATA MODEL

The principle of a Multicarrier transmitter is depicted in Figure 1. The concept is to transform serial data into parallel lower rate inputs that are modulated by orthogonal carriers. History and applications of MCM are reported in [1],[2] and the references therein and are not stated here for conciseness purposes. Assuming a single MCM source

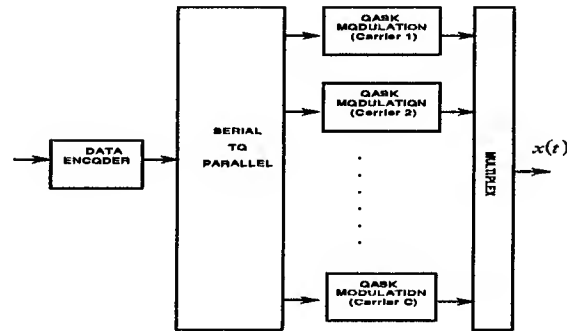


Figure 1: Block diagram of the MCM transmitter.

emitting over  $C$  carriers, the lowpass equivalent transmitted signal is given by

$$x(t) = \sum_{c=1}^C \sum_{k=-\infty}^{\infty} s_c[k] g(t - kT) e^{2\pi j \frac{c}{C} t} \quad (1)$$

where

- $s_c[k]$  is the  $k$ -th symbol conveyed by carrier  $c$ ,
- $\{s_c[k]\}$ ,  $c = 1, \dots, C$  are independent from one carrier to another and identically distributed,
- $g(t)$  is the pulse-shape function,
- $T$  is the symbol duration, and

- the frequency spacing between two successive carriers is  $\frac{1}{T}$ .

In the following, the channel is fading and time varying. However, it is regarded to be stationary within its coherence time. Assuming  $C$  carriers and perfect carrier phase and sampling time recovery, the complex envelope of the lowpass received signal at an  $M$ -element antenna array at time  $t$  can be written as

$$\begin{aligned} \mathbf{y}(t) &= \sum_{c=1}^C \mathbf{y}_c(t) e^{j2\pi \frac{c}{C} t} + \mathbf{z}(t) \\ &= \sum_{c=1}^C \sum_{k=-\infty}^{\infty} s_c[k] \mathbf{h}_c(t - kT) e^{j2\pi \frac{c}{C} t} + \mathbf{z}(t) \quad (2) \end{aligned}$$

where  $\mathbf{h}_c(t) = [h_{c,1}(t) \ h_{c,2}(t) \ \dots \ h_{c,M}(t)]^T$  is the transmission channel associated with the  $c$ -th carrier,  $s_c[k]$  is the  $k$ -th symbol of duration  $T$  conveyed by carrier  $c$  and  $\mathbf{z}(t)$  is the additive white Gaussian noise. The coherence time of the channel is assumed to range over  $K$  symbol periods. The channel  $\mathbf{h}_c(t)$  can be modeled as [3]

$$\mathbf{h}_c(t) = \sum_{q=1}^Q \mathbf{a}(\theta_q) \beta_c(q) g(t - \tau_q) e^{-j2\pi c \frac{\tau_q}{T}} \quad (3)$$

where  $Q$  is the number of paths,  $\theta_q$  and  $\tau_q$  are the  $q$ -th angle of arrival and time delay respectively and  $\beta_c(q)$  is the complex attenuation, which is varying from carrier to carrier.  $\mathbf{a}(\theta_q)$  is the  $(M \times 1)$  vector of the array response to the  $q$ -th path, with  $q = 1, \dots, Q$  and  $g(t)$  is the finite support modulation pulse-shape function. We assume that the array outputs are received in parallel over each carrier after demodulation. The channel length is  $LT$ . We collect  $K$  data samples on each carrier. Using some trivial manipulations, this can be expressed in a  $(M \times K)$ -dimensional matrix form as

$$\mathbf{Y}_c = \mathbf{H}_c \mathbf{S}_c \quad c = 1, \dots, C \quad (4)$$

If the Toeplitz matrix of data symbols  $\mathbf{S}_c$ ,  $c = 1, \dots, C$ , is known from training and  $K \geq M$ , an estimate of the channel samples matrix  $\mathbf{H}_c$  in (3) can be obtained for  $c = 1, \dots, C$ , using least squares. Blind estimation of the channel samples [4, 5] is also possible in case  $\mathbf{S}_c$  is not known in advance. The estimated channel can be given as

$$\hat{\mathbf{H}}_c = \mathbf{H}_c + \mathbf{N}_c \quad (5)$$

where  $\mathbf{N}_c$  is the estimation noise matrix.

Omitting the estimation noise, one can easily show that for each carrier, the terms  $e^{-j2\pi c \frac{\tau_q}{T}}$ ,  $q = 1, \dots, Q$  in equation (5) can be factored out, resulting in

$$\mathbf{H}_c := \mathbf{A}_c \text{diag}[\mathbf{e}_c(\tau)] \mathbf{G} \quad c = 1, \dots, C \quad (6)$$

where the  $(i, j)$ -th element of  $\mathbf{G}$  is defined as

$$\mathbf{G}_{i,j} = g((j-1)T - \tau_i), i = 1, \dots, Q \text{ and } j = 1, \dots, L$$

and

$$\mathbf{A}_c(\theta) = [\beta_1(c) \mathbf{a}(\theta_1) \ \beta_2(c) \mathbf{a}(\theta_2) \ \dots \ \beta_Q(c) \mathbf{a}(\theta_Q)] \quad (7)$$

with

$$\begin{aligned} \mathbf{e}_c(\tau) &= \begin{bmatrix} 1 \\ e^{-j2\pi \frac{c}{C} \tau_1} \\ \vdots \\ e^{-j2\pi \frac{c}{C} \tau_Q} \end{bmatrix} \quad c = 1, \dots, C \\ \theta &= [\theta_1 \ \theta_2 \ \dots \ \theta_Q]^T \quad (8) \end{aligned}$$

and

$$\tau = [\tau_1 \ \tau_2 \ \dots \ \tau_Q]^T \quad (9)$$

If we stack all the matrices  $\mathbf{H}_c$  corresponding to all the  $C$  carriers, we will obtain a large  $(MC \times L)$ -dimensional matrix  $\mathcal{H}$  whose structure is given by

$$\begin{aligned} \mathcal{H} &= \begin{bmatrix} \mathbf{H}_1 \\ \mathbf{H}_2 \\ \vdots \\ \mathbf{H}_C \end{bmatrix} \\ &:= \mathbf{U}(\theta, \tau) \mathbf{G} \quad (10) \end{aligned}$$

where

$$\mathbf{U}(\theta, \tau) = \begin{bmatrix} \mathbf{A}_1(\theta) \text{diag}[\mathbf{e}_1(\tau)] \\ \mathbf{A}_2(\theta) \text{diag}[\mathbf{e}_2(\tau)] \\ \vdots \\ \mathbf{A}_C(\theta) \text{diag}[\mathbf{e}_C(\tau)] \end{bmatrix} \quad (11)$$

Finally, we include the channel estimation noise matrix  $\mathcal{N}$ , which is appropriately defined in accordance with (5) and (10). Therefore, the model in (10) becomes

$$\hat{\mathcal{H}} = \mathbf{U}(\theta, \tau) \mathbf{G} + \mathcal{N} \quad (12)$$

If we consider that the delay spread of the channel is  $T_m$  (expressed in terms of the symbol period  $T$ ), then the coherence bandwidth of the channel is roughly the inverse of  $T_m$ , i.e.,

$$B_{coh} = \frac{1}{T_m} = \frac{1}{nT}$$

The frequency separation between carriers in the MultiCarrier system is given by  $\Delta f = \frac{1}{T}$ . All the carriers that lie within a frequency interval equal to the channel coherence bandwidth can be seen as identically attenuated. Therefore, it is reasonable to assume that the number of carriers being attenuated equally is  $\mu = \lfloor \frac{B_{coh}}{\Delta f} \rfloor = \lfloor \frac{1}{n} \rfloor$ , where  $\lfloor \cdot \rfloor$  denotes the integer part. Under this condition, the number of  $\mu$ -carrier sets that share the same attenuation coefficients is obviously  $m = nC$ .

If we consider only the first  $m\mu$  carriers in the derivation of the MC-JADE model (10) ( $m\mu$  is at most equal to  $C$ ), we will obtain a reduced MC-JADE model ( $Mm\mu \times L$ ) satisfying the following factorization

$$\begin{aligned} \hat{\mathcal{H}}_{m\mu} &= \mathbf{U}_{m\mu}(\theta, \tau) \mathbf{G} + \mathcal{N}_{m\mu} \\ &:= \begin{bmatrix} \mathcal{F}_1(\tau) \circ \mathbf{A}_1(\theta) \\ \mathcal{F}_2(\tau) \circ \mathbf{A}_2(\theta) \\ \vdots \\ \mathcal{F}_m(\tau) \circ \mathbf{A}_m(\theta) \end{bmatrix} \mathbf{G} + \mathcal{N}_{m\mu} \quad (13) \end{aligned}$$



where

$$\mathbf{A}_i(\theta) = \begin{bmatrix} \beta_1(i)\mathbf{a}(\theta_1) & \beta_2(i)\mathbf{a}(\theta_2) & \dots & \beta_Q(i)\mathbf{a}(\theta_Q) \end{bmatrix}$$

$$\mathcal{F}_i(\tau) = \begin{bmatrix} \phi_1^i & \phi_2^i & \dots & \phi_Q^i \\ \phi_1^{i+1} & \phi_2^{i+1} & \dots & \phi_Q^{i+1} \\ \vdots & \vdots & \ddots & \vdots \\ \phi_1^{i+\mu-1} & \phi_2^{i+\mu-1} & \dots & \phi_Q^{i+\mu-1} \end{bmatrix}$$

and

$$\phi_q = e^{-j2\pi\frac{\tau q}{T}} \quad (14)$$

$\circ$  denotes Khatri-Rao product, i.e., columnwise Kronecker Product.

### 3. THE ALGORITHM: MC-JADE-ESPRIT

If the array is Uniform Linear (ULA) or has an ESPRIT doublet structure, then the angles and delays can be estimated jointly in closed-form using an ESPRIT-like method. For the ULA geometry, the steering vector  $\mathbf{a}(\theta_q)$  will be given by

$$\mathbf{a}(\theta_q) = \begin{bmatrix} 1 & \psi_q & \dots & \psi_q^{M-1} \end{bmatrix}^T \quad (15)$$

where

$$\psi_q = e^{j2\pi\Delta \sin \theta_q} \quad (16)$$

and  $\Delta$  is the array sensor spacing in wavelengths.

With the parameter definitions (16) and (14), it is more appropriate to rewrite (13) as

$$\tilde{\mathcal{H}}_{m\mu} = \mathbf{U}(\psi, \phi)\mathbf{G} + \mathcal{N}_{m\mu} \quad (17)$$

where

$$\psi = \begin{bmatrix} \psi_1 & \psi_2 & \dots & \psi_q \end{bmatrix}^T \quad (18)$$

and

$$\phi = \begin{bmatrix} \phi_1 & \phi_2 & \dots & \phi_q \end{bmatrix}^T \quad (19)$$

Estimation of the channel subspace and its dimension is equivalent to finding a basis  $\mathbf{E}$  of the column span of the data matrix  $\tilde{\mathcal{H}}_{m\mu}$  and estimating of the parameters  $\psi$  and  $\phi$  reduces to jointly diagonalize the matrices  $\mathcal{E}_\psi^\dagger \mathcal{E}'_\psi$  and  $\mathcal{E}_\phi^\dagger \mathcal{E}'_\phi$ , where

$$\begin{cases} \mathcal{E}_\psi = \mathbf{J}_\psi \mathbf{E} \\ \mathcal{E}'_\psi = \mathbf{J}'_\psi \mathbf{E} \end{cases} \quad (20)$$

and

$$\begin{cases} \mathcal{E}_\phi = \mathbf{J}_\phi \mathbf{E} \\ \mathcal{E}'_\phi = \mathbf{J}'_\phi \mathbf{E} \end{cases} \quad (21)$$

where

$$\begin{cases} \mathbf{J}_\psi = \mathbf{I}_{m\mu} \otimes \begin{bmatrix} \mathbf{I}_{M-1} & \mathbf{0}_{(M-1,1)} \\ \mathbf{0}_{(M-1,1)} & \mathbf{I}_{M-1} \end{bmatrix} \\ \mathbf{J}'_\psi = \mathbf{I}_{m\mu} \otimes \begin{bmatrix} \mathbf{I}_{M-1} & \mathbf{0}_{(M-1,1)} \\ \mathbf{0}_{(M-1,1)} & \mathbf{I}_{M-1} \end{bmatrix} \end{cases} \quad (22)$$

and

$$\begin{cases} \mathbf{J}_\phi = \mathbf{I}_m \otimes \begin{bmatrix} \mathbf{I}_{M(\mu-1)} & \mathbf{0}_{(M(\mu-1)),M} \\ \mathbf{0}_{(M(\mu-1)),M} & \mathbf{I}_{M(\mu-1)} \end{bmatrix} \\ \mathbf{J}'_\phi = \mathbf{I}_m \otimes \begin{bmatrix} \mathbf{I}_{M(\mu-1)} & \mathbf{0}_{(M(\mu-1)),M} \\ \mathbf{0}_{(M(\mu-1)),M} & \mathbf{I}_{M(\mu-1)} \end{bmatrix} \end{cases} \quad (23)$$

are the appropriate selection matrices (see [6],[7] and [8] for details of JADE),  $\otimes$  denotes Kronecker product,  $\mathbf{I}_i$  is an  $i$ -dimensional identity matrix and  $\mathbf{0}_{i,j}$  is a  $(i \times j)$ -dimensional matrix of zero elements.

Details of the joint diagonalization are provided in [7] and the references therein. The correct pairing between the  $\psi$ 's and the  $\phi$ 's is guaranteed by the fact that matrices share common eigenvectors.

If the pulse-shape function is assumed to be known, the complex attenuation coefficients can be linearly estimated using least-squares, by processing the channel samples over each carrier separately.

### 4. IDENTIFIABILITY

The parameter identifiability requires to have the  $(Mm\mu \times L)$ -dimensional data matrices  $\tilde{\mathcal{H}}_i$  of rank  $Q$ , with  $Q < Mm\mu$  and  $Q \leq L$ . This means that  $\mathbf{U}_i(\theta, \tau)$  must have strictly more rows than columns and be of full column rank, and  $\mathbf{G}$  must have more columns than rows and be of full row rank. The full rank condition on  $\mathbf{G}$  together with the channel factorization (6) imply that all the delays must be distinct. If two paths have the same TDOA's, the rank of  $\tilde{\mathcal{H}}$  becomes  $Q - 1$  and the corresponding angles cannot be identified correctly. In this case, "spatial smoothing" [7] can provide the solution [6],[7] by performing data extension of the channel over each carrier in such a way to keep rank  $\tilde{\mathcal{H}}_i$  equal to the number of paths  $Q$ . In order to allow selection of the received data (13), there must be at least a pair of sensors, i.e.,  $M \geq 2$ , and the coherence bandwidth to carrier frequency-spacing ratio must be at least 2 : 1, i.e.,  $\frac{B_{coh}}{\Delta f} \geq 2$  or  $\mu \geq 2$ . The last requirement can be satisfied by appropriately increasing the number of carriers.

### 5. SIMULATIONS

The following simulation results illustrate performance of MC-JADE-ESPRIT. In all the experiments, the estimation Mean Square Error (MSE) is averaged over 500 Monte Carlo runs of the algorithm and compared against the Cramer-Rao Bound (CRB) which is derived for the model (13) in the Appendix. In the figures corresponding to the experiments, the MSE is plotted using a full line whereas the CRB is shown by a dotted line.

#### 5.1. Basic performance of MC-JADE-ESPRIT

We consider an antenna of  $M = 2$  elements, spaced at half wavelength. The number of paths is  $Q = 3$  with parameters  $\theta = [-15^\circ \ 0^\circ \ 25^\circ]^T$ ,  $\tau = [0 \ 0.078 \ 0.234]^T T$  and the path fadings being generated from a complex zero-mean Gaussian distribution with variance [0.4 0.3 0.3]. The channel length is half the symbol period  $T$ , which is normalized to  $T = 1$ . The pulse-shape function is a raised cosine with 0.25 roll-off factor.  $C = 64$ , with  $\mu = 8$ . The employed joint diagonalization method is method "Q" as

it is referred to in [7]. Fig. 2 shows the effect of the noise power on the MSE of the estimated DOA's and TDOA's. At high noise powers, the estimation is strongly sensitive to the channel estimation noise and is erroneous. As the noise effect decreases, the difference with the CRB is about 2 to 3 dB.

### 5.2. Comparison with SI-JADE

For the same setting, we plot the CRB relative to the parameter estimation over the first carrier, using SI-JADE [7], against the noise power. The stacking parameter as defined in [7] is taken to be  $m_1 = 5$ . The CRB of SI-JADE is plotted in Fig. 2 using a dashed-line. Here, for low estimation noise powers, the parameter MSE of MC-JADE-ESPRIT is smaller than the CRB of SI-JADE. The greater estimation precision for MC-JADE-ESPRIT is mainly due to the larger amount of information involved.

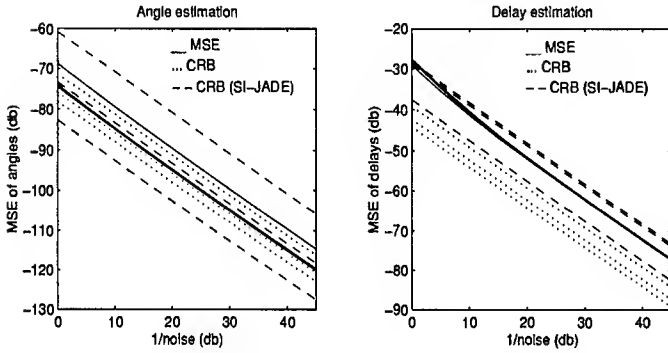


Figure 2: Basic performance of MC-JADE-ESPRIT.

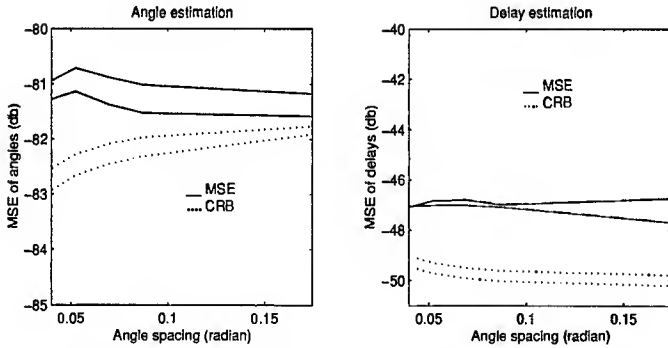


Figure 3: Spatial resolution of MC-JADE-ESPRIT.

### 5.3. Resolution of the Algorithm

We set the number of paths to  $Q = 2$ , with the estimation noise power being fixed at -20 dB. All the other parameters are kept the same. In Fig. 3, as it is expected, it is shown that estimation accuracy improves with well separated angles, else estimation is dependent on noise power. The effect of delay spacing on the angle and delay estimation is shown

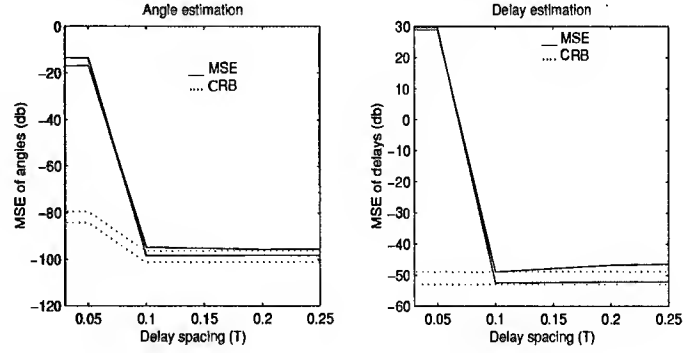


Figure 4: Temporal resolution of MC-JADE-ESPRIT.

on Fig. 4. It is clear that for small delay spacing, ambiguity occurs and the full rank condition on the pulse-shape function matrix is no more satisfied, yielding an erroneous estimation. Here, no spatial smoothing is applied. For well separated delays, estimation is seen to depend only on the noise power.

## 6. CONCLUSION

Advantage of the algorithm is that it takes into account the available frequency diversity provided by the multiple carriers and processes data in a single batch. However, estimation of the channel impulse response is prerequisite to the application of the algorithm, which makes its performance suboptimal and sensitive to the estimation noise.

## Appendix The Cramer Rao Bound

The CRB for the joint problem (13) can be derived as follows:

Let us define the parameter vector as

$$[\sigma_N^2 \mathbf{g}^T(1) \dots \mathbf{g}^T(L) \boldsymbol{\eta}^T]$$

where

$$\boldsymbol{\eta} := [\Re\{\beta^T(1)\} \Im\{\beta^T(1)\} \dots \Re\{\beta^T(m)\} \Im\{\beta^T(m)\} \theta^T \tau^T]^T$$

and  $\Re\{\cdot\}$  and  $\Im\{\cdot\}$  denote the real and imaginary parts respectively. In our case, vectors  $\mathbf{g}(i), i = 1, \dots, L$ , which are the columns of matrix  $\mathbf{G}$  in (13), are deterministic but unknown. The data are the channel estimates  $\tilde{\mathbf{H}}_{m\mu}$ . These data are corrupted by the estimation noise

$$\mathcal{N}_{m\mu} := [\mathbf{n}(1) \mathbf{n}(2) \dots \mathbf{n}(L)]$$

where  $\mathbf{n}(i), i = 1, \dots, L$  are complex, stationary, zero-mean Gaussian random processes that are temporally uncorrelated. It follows that the data  $\tilde{\mathbf{H}}_{m\mu}$  are also uncorrelated Gaussian random processes. The likelihood function of the data is

$$\mathcal{L}\{\tilde{\mathbf{H}}_{m\mu}\} = \frac{1}{(2\pi)^{Mm\mu L} \left(\frac{\sigma_N^2}{2}\right)^{Mm\mu L}} \times$$

$$\times \exp \left\{ -\frac{1}{\sigma_N^2} \sum_{i=1}^L \mathbf{n}^*(i) \mathbf{n}(i) \right\} \quad (24)$$

and the corresponding loglikelihood function is

$$\Lambda = \ln \mathcal{L} = \text{const} - Mm\mu L \ln \sigma_N^2 - \frac{1}{\sigma_N^2} \sum_{i=1}^L \mathbf{n}^*(i) \mathbf{n}(i) \quad (25)$$

where  $*$  denotes complex conjugate transpose. The derivatives of the loglikelihood function  $\Lambda$  with respect to the unknown parameters can be obtained using results of [9],[6],[10], as

$$\begin{aligned} \frac{\partial \Lambda}{\partial (\sigma_N^2)} &= -\frac{Mm\mu L}{\sigma_N^2} + \frac{1}{\sigma_N^4} \sum_{i=1}^L \mathbf{n}^*(i) \mathbf{n}(i) \\ \frac{\partial \Lambda}{\partial (\mathbf{g}(i))} &= \frac{2}{\sigma_N^2} \Re[\mathbf{U}^* \mathbf{n}(i)] \\ \frac{\partial \Lambda}{\partial \eta} &= \frac{2}{\sigma_N^2} \sum_{i=1}^L \Re\{\mathbf{g}(i) \mathbf{D}^* \mathbf{n}(i)\} \end{aligned}$$

with  $\mathbf{U} = \mathbf{U}(\theta, \tau)$ , and

$$\begin{aligned} \mathbf{D} &:= [\mathbf{D}_\beta \ \mathbf{D}_\theta \ \mathbf{D}_\tau] \quad (Mm\mu \times 2(m+1)Q) \\ \mathbf{D}_\beta &:= [\mathbf{D}_{\Re\{\beta(1)\}} \ \mathbf{D}_{\Im\{\beta(1)\}} \ \dots \ \mathbf{D}_{\Re\{\beta(m)\}} \ \mathbf{D}_{\Im\{\beta(m)\}}] \\ \mathbf{D}_{\Re\{\beta(i)\}} &:= \begin{bmatrix} \frac{\partial \mathbf{U}}{\partial \Re\{\beta(i)_1\}} & \dots & \frac{\partial \mathbf{U}}{\partial \Re\{\beta(i)_q\}} \end{bmatrix} \\ \mathbf{D}_{\Im\{\beta(i)\}} &:= \begin{bmatrix} \frac{\partial \mathbf{U}}{\partial \Im\{\beta(i)_1\}} & \dots & \frac{\partial \mathbf{U}}{\partial \Im\{\beta(i)_q\}} \end{bmatrix} \\ \mathbf{D}_\theta &:= \begin{bmatrix} \frac{\partial \mathbf{U}}{\partial \theta_1} & \dots & \frac{\partial \mathbf{U}}{\partial \theta_q} \end{bmatrix} \\ \mathbf{D}_\tau &:= \begin{bmatrix} \frac{\partial \mathbf{U}}{\partial \tau_1} & \dots & \frac{\partial \mathbf{U}}{\partial \tau_q} \end{bmatrix} \\ \mathbf{g}(i) &:= \mathbf{I}_{2(m+1)} \otimes \mathbf{g}(i) \end{aligned}$$

Using results of [9],[6], we get

$$\begin{aligned} E \left[ \left( \frac{\partial \Lambda}{\partial (\sigma_N^2)} \right)^2 \right] &= \frac{Mm\mu L}{\sigma_N^4} \\ E \left[ \left( \frac{\partial \Lambda}{\partial \mathbf{g}(i)} \right) \left( \frac{\partial \Lambda}{\partial \mathbf{g}(j)} \right)^T \right] &= \frac{2}{\sigma_N^2} \Re[\mathbf{U}^* \mathbf{U}] \delta_{i,j} \\ E \left[ \left( \frac{\partial \Lambda}{\partial \mathbf{g}(i)} \right) \left( \frac{\partial \Lambda}{\partial \eta} \right)^T \right] &= \frac{2}{\sigma_N^2} \Re[\mathbf{U}^* \mathbf{D} \mathbf{g}(i)] \\ E \left[ \left( \frac{\partial \Lambda}{\partial \eta} \right) \left( \frac{\partial \Lambda}{\partial \eta} \right)^T \right] &= \frac{2}{\sigma_N^2} \sum_{i=1}^L \Re[\mathbf{g}^*(i) \mathbf{D}^* \mathbf{D} \mathbf{g}(i)] \end{aligned}$$

The Fisher Information Matrix (FIM) for the parameters is given by  $E(\omega \omega^T)$ , where  $\omega := [\sigma_N^2 \ \mathbf{g}^T(1) \ \dots \ \mathbf{g}^T(L) \ \eta^T]^T$  and the inverse of the CRB matrix for the parameters, after some manipulations, is given by

$$\begin{aligned} \text{CRB}^{-1}(\eta) &= \frac{2}{\sigma_N^2} \sum_{i=1}^L \{ \Re[\mathbf{g}^*(i) \mathbf{D}^* \mathbf{D} \mathbf{g}(i)] \\ &\quad - \Re[\mathbf{U}^* \mathbf{D} \mathbf{g}(i)]^T \Re[\mathbf{U}^* \mathbf{U}]^{-1} \Re[\mathbf{U}^* \mathbf{D} \mathbf{g}(i)] \} \end{aligned}$$

Finally, the CRB matrix for the parameters of interest,  $\text{CRB}(\theta, \tau)$ , is the  $2Q$ -dimensional bottom-right-corner partition matrix of  $\text{CRB}(\eta)$  and the bounds are found by taking the diagonal elements.

## 7. REFERENCES

- [1] J. A. C. Bingham, "Multicarrier Modulation for Data Transmission: An Idea Whose Time Has Come", *IEEE Communications Magazine*, vol. 28, NO. 5, May 1990.
- [2] I. Kalet, "The Multitone Channel", *IEEE Transactions on Communications*, vol. 37, NO. 2, February 1989.
- [3] L. Vandendorpe and O. van de Wiel, "MIMO DFE Equalization for Multitone DS/SS Systems over Multipath Channels", *IEEE Transactions on Communications*, vol. 14, NO.3, April 1996.
- [4] A. Belouchrani and M. G. Amin, "Blind Source Separation Based on Time-Frequency Signal Representations", *IEEE Transactions on Signal Processing*, vol. 46, NO. 11, november 1998.
- [5] K. Abed-Meraim and Y. Hua, "Blind Identification of Multi-Input Multi-Output System Using Minimum Noise Subspace", *IEEE Transactions on Signal Processing*, vol. 45, NO. 1, January 1997.
- [6] M. C. Vanderveen, A. J. van der Veen and A. Paulraj, "Estimation of Multipath Parameters in Wireless Communications", *IEEE Transactions on Signal Processing*, vol. 46, NO.3, March 1998.
- [7] A. J. van der Veen, M. C. Vanderveen, and A. Paulraj, "Joint Angle and Delay Estimation Using Shift-Invariance Techniques" *IEEE Transactions on Signal Processing*, vol. 46, NO.2, February 1998.
- [8] A. J. van der Veen, M. C. Vanderveen and A. Paulraj, "Joint Angle and Delay Estimation Using Shift Invariance Properties" *IEEE Signal Processing Letters*, vol. 4, NO. 5, May 1997.
- [9] P. Stoica and A. Nehorai, "MUSIC, Maximum Likelihood, and the Cramer-Rao Bound", *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 37, NO. 5, May 1989.
- [10] S. M. Kay, "Fundamentals of Statistical Signal Processing: Estimation Theory", Prentice-Hall, 1993.

# OFDM SPECTRAL CHARACTERIZATION: ESTIMATION OF THE BANDWIDTH AND THE NUMBER OF SUB-CARRIERS

Walter AKMOUCHE

CELAR/TCOM/TR - BP 7  
35 174 BRUZ CEDEX - FRANCE  
e-mail: akmouche@celar.fr

Eric KERHERVE, André QUINQUIS

ENSIETA - 2, rue F. VERNY  
29 200 BREST - FRANCE  
quinquis@ensieta.fr

## ABSTRACT

This paper deals with the analysis of modulated signals in a NDA (Non Data Aided) context. Assuming the detection of an OFDM signal, our goal is to estimate the bandwidth and the number of sub-carriers of this signal. First, we propose an algorithm based on wavelet decomposition in order to estimate the bandwidth: bandwidth is correctly estimated in 100 % of the cases with an error lower than 8 % until SNR = 3 dB. Second, we apply the MUSIC algorithm with decision criterion to obtain the number of sub-carriers: the number of carriers can be estimated with an error lower or equal to 9 % in 100 % of the cases until SNR = 10 dB.

## 1. INTRODUCTION

Spectrum survey requires the estimation of the parameters of the received signals. This problem has already been studied in the case of the single-carrier modulations, and has now to cope with new modulations types like OFDM (Orthogonally Frequency Division Multiplexing) which are more and more used (DAB, ADSL,...). In [2], we proposed a method to detect OFDM signals versus linear single-carrier modulated signals. The problem we now want to solve is the estimation of two main parameters of such a signal: the bandwidth and the number of sub-carriers. Using the fact that the power spectral density (PSD) of an OFDM signal has a rectangular shape, we propose to apply a wavelet decomposition to detect the breaking points at the beginning and at the end. Then, we try to determine the number of sub-carriers. Since this number is unknown, AR modelization is impossible. Therefore, the MUSIC algorithm with decision criterion seems to be well suited to solve this problem. In section 2 we give the problem statement. Section 3 is dedicated to the bandwidth estimation of OFDM signals, with performances. In section 4 we give a method to obtain the number of sub-carriers. Section 5 concludes the paper.

## 2. PROBLEM STATEMENT

OFDM is a single carrier multiplexing, and can then be expressed as a sum of single carrier modulated signals:

$$x_m(t) = \sqrt{\frac{P}{N_p}} \sum_k \sum_{n=0}^{N_p-1} c_{n,k} \cdot e^{2i\pi(f_0+n\Delta f)t} \cdot g(t - kT_s) \quad (1)$$

where  $\{c_{n,k}\}$  is the symbol sequence which is assumed to be centered, i.i.d.,  $N_p$  the number of sub-carriers,  $\Delta f$  the frequency offset between carriers,  $g(t)$  the pulse function and  $P$  the power of the signal.  $T_s = T_u + T_g$ ,  $T_u$  is the "useful time" when information is sent,  $T_g$  is the interval guard and  $T_s$  the time of the complete OFDM symbol. We will suppose here that the interval guard is empty. Due to the multiplexing of many single carrier signals, the spectrum of the OFDM signal is quite rectangular (Fig. 1). We assume to receive the complex signal  $r(t) = x(t) + b(t)$  where  $x(t)$  is the OFDM baseband signal (with possible frequency and time offsets) and  $b(t)$  is a complex white gaussian noise.

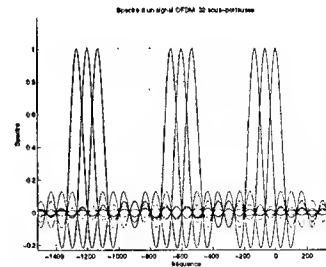


Figure 1: Spectrum amplitude of OFDM signal with 32 carriers.

## 3. BANDWIDTH ESTIMATION

### 3.1. Continuous wavelet decomposition (CWT)

From a signal point of view, wavelets consist of a linear decomposition of a signal on a given waveform translated in time and dilated or compressed in time [1]. In the frequency domain, wavelet analysis is closely related to fil-

tering the data through a bank of filters having constant surtension coefficients. The continuous wavelet transform (CWT) maps a one-dimensional analog signal called  $s(t)$  to a set of wavelet coefficients which vary continuously over time  $b$  and scale  $a$ :

$$W(a, b) = a^{-1/2} \cdot \int_{-\infty}^{+\infty} \psi^*\left(\frac{t-b}{a}\right) \cdot s(t) \cdot dt$$

where  $W(a, b)$  signifies "Wavelet Transform".  $\psi(t)$  is the wavelet used in the decomposition. Equivalently the CWT can be expressed as:

$$W(a, b) = a^{1/2} \cdot \int_{-\infty}^{+\infty} \bar{\psi}^*(a\nu) \cdot S(\nu) \cdot e^{2i\pi\nu b} \cdot d\nu$$

with  $\bar{\psi}(\nu)$  and  $S(\nu)$  the Fourier transforms of  $\psi(t)$  and  $s(t)$  respectively. Wavelets must satisfy some restrictions [1], the most important ones are integrability and square integrability. Consequently, this condition implies that if  $\bar{\psi}(\nu)$  is a smooth function in the neighborhood of the frequency origin then  $\psi(0) = 0$ , which means that  $\psi(t)$  has no DC component. Other assumptions about wavelets can be made for convenience. One such requirement is that  $\bar{\psi}(\nu) = 0$ , for  $\nu < 0$ . It is also convenient to assume that  $\bar{\psi}(\nu)$  is real for  $\nu > 0$ . The wavelet functions  $\psi(\frac{t-b}{a})$  are used to band-pass filter the signal. This can be seen as a kind of time-varying spectral analysis in which scale  $a$  plays the role of a local frequency. As  $a$  increases, wavelets are stretched and analyze low frequencies, while for small  $a$ , contracted wavelets analyze high frequencies. The parameter  $b$  varying in time controls the desired temporal location. The scalar product corresponds to the signal measurement  $s(t)$  in the space drawn by all the dilated or contracted figures of unique function  $\psi$ . In order to analyze, the dilation parameter  $a$  is given an initial large value (e.g. 1.0) and is then decreased in regular increments to examine the signal in more detail. We can write equivalently that the wavelet filter function considers successively narrow section of the signal spectrum  $S(\nu)$ . Since spectral properties are frequently better displayed on a logarithmic frequency scale, it is convenient to write  $a = 2^{-u}$ . With this notation integral increments in  $u$  result in octave increments of  $a$ . Note that a small  $a$  (i.e. large  $u$ ) corresponds to high frequencies. A small  $u$  corresponds to an analysis of the large scale features of  $s(t)$ , and as  $u$  increases, finer details of the signal come into focus. The function  $\psi(t)$  is the basic unshifted and undilated wavelet. It may be chosen to answer the needs [5]. For example, in our case,  $\psi(t) = e^{-\frac{t^2}{2} + jmt}$  is the Morlet wavelet. An important property of this basic wavelet is that it is concentrated in the time and frequency domains. This means that the time-bandwidth product is as small as possible. To satisfy  $\bar{\psi}(0) = 0$ , one must add a correction term, but if  $m > 5$ , this correction term is negligibly small and can be omitted. One problem of practical interests for engineers is detection of abnormal features. Generally, we have to use a discretization procedure since we consider

digital data. This discretization procedure consists in a high resolution digitalization of the generating wavelet in the time domain, truncated on its sides in order to have a finite extent. Then, the wavelet coefficients  $C_{j,k}$  of the time-frequency decomposition are obtained by a correlation in the time-domain of the interpolated digitized wavelets  $\psi_{j,k}$  with the discrete signal  $s(n)$  for different values of the dilation factor  $2^j$  and of the time shift  $k$ . This approach presents some drawbacks such as the edge effects due to the correlation of a finite duration signal with a truncated infinite wavelet, the numerical approximations due to truncature,...

### 3.2. Bandwidth estimation method

The beginning and the end of the PSD of an OFDM signal, called  $R(f)$ , are breaking points and can be easily detected by using a wavelet decomposition [4]. We decide to choose the Morlet wavelet for analyzing the PSD signal and obtain the scalogram figure of the PSD (Fig. 2). Nevertheless, we have to admit that the esti-

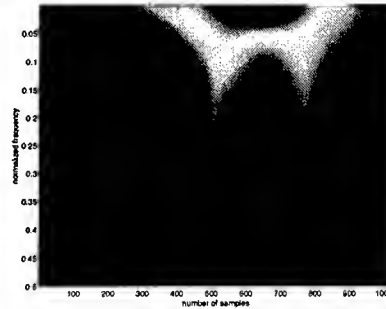


Figure 2: Scalogram of the PSD of the received signal  $r(t)$ . 1024 samples. SNR = 3 dB.

mation is purely visual. For that reason, we decide to project the resulted scalogram to obtain its frequency marginal. Because wavelet analysis is a constant  $\Delta f/f$  transformation, we have to make the sum of energy in a cone, instead of summing energy of column as in the case of a bilinear time-frequency transformation. Moreover, we can not be sure that the wavelet has the same energy in each time-frequency logon. Consequently, we propose to calculate the scalogram of the Dirac distribution which has a cone shape and specifically characterizes breaking points. Considering this scalogram, it becomes easy to conserve only points with enough energy (i.e. more energy that a given percentage of the total energy of the signal) and then to form a mask of description. We then obtain the bandwidth estimation algorithm:

1. Apply the Dirac mask on the scalogram of the studied PSD signal  $R(f)$  for each frequency localization.
2. Calculate the sum of the energy, which gives the frequency marginal of the scalogram .
3. Search for the two extrema located in the beginning and the end of the bandwidth.

Two options are possible to calculate the energy of the scalogram of  $R(f)$  in the cone of the Dirac mask. First, we can use a binary mask, which means that energy is equal to "1" if the point belongs to the cone, "0" otherwise. The second solution consists in using a weighted Dirac mask which gives the real energy of each logon after thresholding. We show in Fig. 3 that the second solution leads to the right frequency marginal.

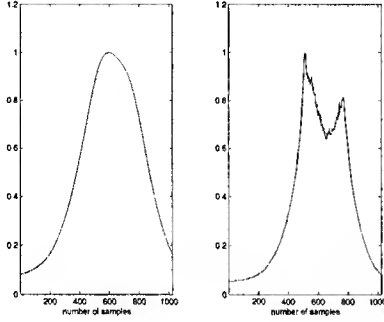


Figure 3: Frequency marginals of the scalogram in the case of binary and weighted Dirac mask.

### 3.3. Results and performance

We apply the proposed algorithm to 10,000 trials of simulated OFDM signals. These signals are generated with 4096 samples, with 4 samples per symbol. The PSD is evaluated by using 1024 points. We have simulated exactly the binary random sequence for SNR equal to 10, 5, 3 and 0 dB. Moreover, we have studied the effects of bad synchronization by considering time and frequency offsets (time offset is smaller  $T_u$  and frequency offset can not exceed 5 % of the bandwidth of the signal).

#### 3.3.1. Noise influence

Fig. 4 shows the results of bandwidth estimation for different SNR. The proposed algorithm permits to determine the bandwidth with a precision lower than 4 % for 97 % of the signals when  $SNR = 3$  dB. But we can observe a strong degradation of the performance as SNR goes to 0 dB.

#### 3.3.2. Time and frequency offset influence

Fig. 5 shows results obtained for different SNR in the case where the frequency offset  $\delta f_0$  is non zero. The new scalogram is quite a translated version of the original scalogram with length  $\delta f_0$ . Consequently, the bandwidth remains the same and the performances are still good. The time offset  $\delta t_0$  is equivalent to a new phase for the signal. Since we evaluate its PSD, phase has not influence anymore and then the performances are strictly the same as in the case  $\delta t_0 = 0$ .

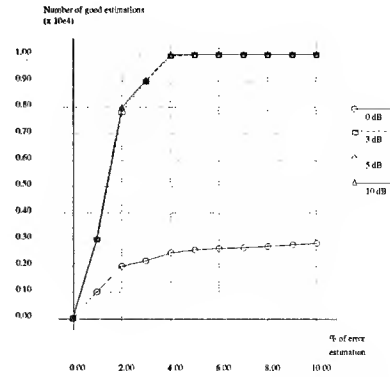


Figure 4: Noise influence: estimation performance for different SNR, no time or frequency offsets.

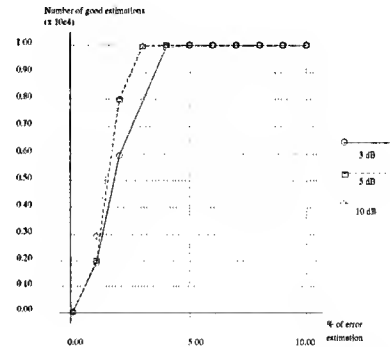


Figure 5: frequency offset influence: estimation performance for different SNR, no time offset.

#### 3.3.3. Conclusion concerning the method

The proposed method is efficient until  $SNR = 3$  dB, even in the case of time or frequency offset. By using the PSD of the received signal, all phase perturbations can be removed. Until  $SNR = 3$  dB, we can conclude that the bandwidth is correctly estimated in 100 % of the cases with an error lower than 8 %.

## 4. ESTIMATION OF THE NUMBER OF SUB-CARRIERS

### 4.1. Theoretical covariance matrix

In this problem, we are receiving one signal which is made of  $N_p$  components. Then, we compute the coefficients of the covariance matrix called  $R$ . For each time-delay  $\tau_n$  in the interval  $[0 ; N_p - 1]$ , the covariance term can be expressed by:

$$r(\tau_n) = \frac{1}{N_p - \tau_n} \cdot \sum_{q=\tau_n+1}^{N_e} x(q) \cdot x^*(q - \tau_n) \quad (2)$$

where  $N_e$  denotes the number of samples of the received signal. Moreover we can notice that the estimator is a

non-biased estimator. In the case where  $\tau_n = 0$ , we have:

$$r(0) = \frac{1}{N_p} \cdot \sum_{q=1}^{N_e} x(q) \cdot x^*(q)$$

$$\sum_{q=1}^{N_e} |x(q)|^2 + \sigma_b^2$$

where  $\sigma_b^2$  is the variance of noise. If  $\tau_n \neq 0$ , we have:

$$r(\tau_n) = \frac{1}{N_p - \tau_n} \cdot \sum_{q=1+\tau_n}^{N_e} x(q) \cdot x^*(q - \tau_n)$$

$$\sum_{q=1+\tau_n}^{N_e} x^2(q) \cdot e^{2i\pi\Delta f\tau_n}$$

and then we consider  $\omega_n = 2\pi\Delta f\tau_n$ , depending on  $\tau_n$ .

We can then form the covariance matrix as:

$$\mathbf{R} = \begin{pmatrix} r(0) & r(1) & \dots & r(N_p - 2) & r(N_p - 1) \\ r^*(1) & r(0) & \dots & r(N_p - 3) & r(N_p - 2) \\ \dots & \dots & \dots & \dots & \dots \\ r^*(N_p - 1) & r^*(N_p - 2) & \dots & r^*(1) & r(0) \end{pmatrix}$$

Considering the value of  $r(\tau_n)$  in the case where  $\tau_n = 0$  or  $\tau_n \neq 0$ , this matrix can also be written as:

$$\mathbf{R} = \begin{pmatrix} \sum_{q=1}^{N_e} x_q^2 + \sigma_b^2 & \sum_{q=1}^{N_e} x_q^2 \cdot e^{i\omega_n} & \dots & \sum_{q=1}^{N_e} x_q^2 \cdot e^{i(N_p-1)\omega_n} \\ \sum_{q=1}^{N_e} x_q^2 \cdot e^{i\omega_n} & \dots & \dots & \sum_{q=1}^{N_e} x_q^2 \cdot e^{i(N_p-2)\omega_n} \\ \dots & \dots & \dots & \dots \\ \sum_{q=1}^{N_e} x_q^2 \cdot e^{i(N_p-1)\omega_n} & \dots & \dots & \sum_{q=1}^{N_e} x_q^2 + \sigma_b^2 \end{pmatrix}$$

This matrix is a symmetrical matrix and its form is the same as in the cases for which MUSIC algorithm is used. Then it can be diagonalized by using eigenvalue decomposition [6]. After the diagonalization process, we know that the autocorrelation matrix becomes:

$$\mathbf{R}_d = \begin{pmatrix} \lambda_1 & 0 & \dots & \dots & \dots & \dots & 0 \\ 0 & \lambda_2 & 0 & \dots & \dots & \dots & \dots \\ \dots & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & \dots & \lambda_{N_p} & 0 & \dots & \dots \\ 0 & \dots & \dots & 0 & \sigma_b^2 & 0 & \dots \\ 0 & \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & \dots & \dots & \dots & \dots & 0 & \sigma_b^2 \end{pmatrix}$$

where  $\lambda_1, \lambda_2, \dots, \lambda_{N_p}$  are the eigenvalues due to the contribution of the useful signal plus noise. Normally,  $\lambda_i > \sigma_b^2, \forall i \in \{1, 2, \dots, N_p\}$ . We can notice that the matrix contains  $N_p$  eigenvalues which are bigger than the noise variance, and then that the number of sub-carriers can be deduced.

Many solutions are possible to determine which values are due to the contribution of the sub-carriers. As the channel has been surveyed before the signal started, we can assume that the variance of noise has been estimated, with of course incertitude. A second solution is to represent the eigenvalues on a same diagram by increasing value order and to detect a breaking point. But, in the

case of fading, the contributions of some sub-carriers become lower and the breaking point is impossible to find. Another solution is to use a decision criterion: Akaike's or Rissanen's criterion. Akaike's criterion is more suited since it tends to overestimate the number of sources if the signal is oversampled, which could be helpful in the case of fadings. Moreover, this method is efficient only if there is two noise contributions (at least). That is that the number of correlation terms must be at least equal to  $(N_p + 1)$ . The problem is that  $N_p$  is unknown and has to be estimated. The proposed solution is to start the algorithm with an *a priori* number of sub-carrier and to iterate this process until one eigenvalue corresponding to noise or a breaking point appears.

## 4.2. Proposed algorithm

The first algorithm we propose is the following:

1. Fixe *a priori* the size of the matrix:  $N_e$ .
2. Using equation 2, compute the  $N_e$  autocorrelation terms and form correlation matrix.
3. Diagonalize the matrix and apply Akaike's criterion. If the number of sub-space (*i.e.* of sub-carriers) is equal to  $N_e$ , go to step 1 and do  $N_e \leftarrow 2 \cdot N_e$ .

## 4.3. Results

We apply the proposed algorithms to simulated OFDM signals. We simulate 10,000 OFDM signals using 10,000 trials to generate the corresponding symbols. Each signal is generated with 50,000 samples normally and contains 64 sub-carriers. The frequency offset is limited to 10% of the bandwidth of the signal. The channel is the urban channel (COST 207) in order to compare decision criterions. We apply MUSIC algorithm with Akaike's criterion (except in figure 8).

### 4.3.1. Noise influence

In the first case, we are looking for noise influence. We generate OFDM signals for different signal-to-noise ratios (20, 10 and 5 dB). We can notice on Fig. 6 that until 10 dB performances are quite good, but become poor for 5 dB and less. Then, we study the influence of the number of signal samples since we use estimators of autocorrelation terms. SNR is fixed to 20 dB, and the signals are tested with respectively 50,000, 40,000 and 30,000 samples. As forecasted, the performances decrease with the number of samples. Nevertheless, 50,000 samples are enough to obtain good performances (Fig. 7). Lastly, we compare Rissanen's and Akaike's criterion in the case of a signal with 50,000 samples and SNR = 20 dB and 10 dB.

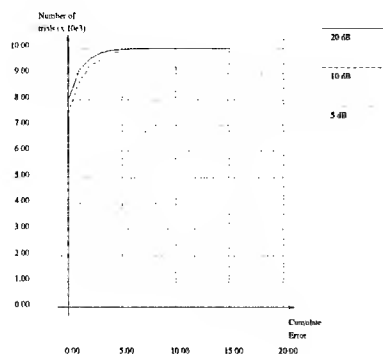


Figure 6: Noise influence in the estimation of the number of sub-carriers.

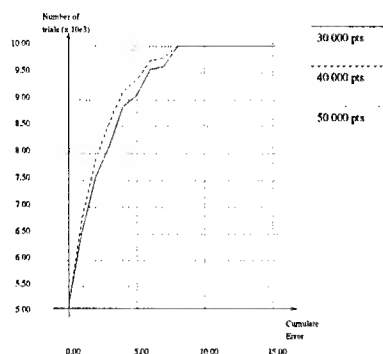


Figure 7: Influence of the number of points in the estimation of the number of sub-carriers. SNR=20 dB.

Since it tends to overestimate the dimension of the signal sub-space, Akaike's criterion is quite better than the Rissanen's one (Fig. 8).

#### 4.4. Conclusion concerning the method.

This method is quite efficient to estimate the number of sub-carriers until SNR = 10 dB and for 50,000 samples (that means about 1500 OFDM symbols). Akaike's criterion is more appropriated than Rissanen's one, but we should test the "Minimum Description Length" criterion.

## 5. CONCLUSION

The proposed methods to estimate the bandwidth and the number of sub-carriers are quite efficient for a few samples and low SNR (lower than 10 dB). Concerning the bandwidth estimation, we obtain a correct estimation in 100 % of the cases with an error lower than 8 % until SNR = 3 dB. Concerning the estimation of the number of sub-carriers, we obtain a correct estimation in 100 % of the cases with an error lower than 9 % until SNR = 10 dB. The performances can be improved using denoising algorithms [3] and compared with time-domain methods that we are currently developing [4]. This work completes our detection algorithm and can be used for coming ap-

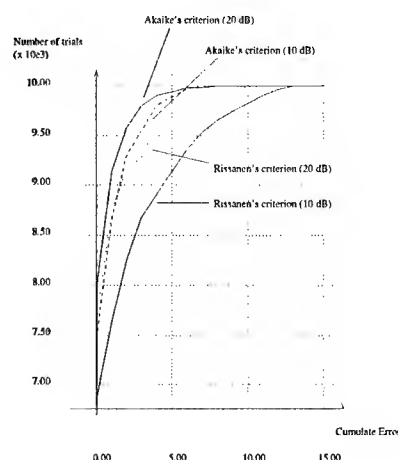


Figure 8: Influence of the decision criterion on the estimation of the number of sub-carriers. 10,000 trials, 50,000 samples, SNR=20 and 10 dB, urban channel (COST 207).

plications of synchronization and equalization.

## REFERENCES

- [1] A. Cohen "Ondelettes et traitement numérique du signal" ed. MASSON, 1992, 205 pages.
- [2] W. Akmouche "Detection of multi-carrier modulations using 4th-order cumulants." *Proc. of the MILCOM*, session 15, Atlantic City, 01-03/11/1999.
- [3] E. Kerherve, W. Akmouche, A. Quinquis "Wavelet and noise reduction: application to the time features estimation for OFDM signals." *Proc. of the ICSPAT*, Orlando (Florida), USA, Oct. 1999.
- [4] W. Akmouche, E. Kerherve, A. Quinquis "Estimation of OFDM signal parameters: time parameters." *submitted to Globecom 2000*, Nov. 2000.
- [4] J.-C. Pesquet, H. Krim, H. Carfantan, J. G. Proakis "Estimation of noisy signals using time-invariant wavelet packets" *Proc. of the IEEE*, 1993, pp. 31-34.
- [5] A. Teolis "Computational signal processing with wavelets" *Proc. of the IEEE*, 1993, pp. 31-34.
- [6] E. H. Attia "Efficient computation of the MUSIC algorithm as applied to a low-angle elevation estimation problem in a severe multipath environment" Ed. Birkhauser, 1998, 324 pages.



# BLIND SOURCE SEPARATION OF NONSTATIONARY CONVOLUTIVELY MIXED SIGNALS

Brian S. Krongold and Douglas L. Jones

Department of Electrical and Computer Engineering & Coordinated Science Laboratory  
University of Illinois at Urbana-Champaign  
Urbana, IL 61801

## ABSTRACT

Many algorithms for blind source separation (BSS) have been introduced in the past few years, most of which assume statistically stationary sources as well as instantaneous mixtures of signals. In many applications, such as separation of speech or fading communications signals, the sources are nonstationary. Furthermore, the source signals may undergo convolutive (or dynamic) linear mixing, and a more complex BSS algorithm is required to achieve better source separation. We present a new BSS algorithm for separating linear convolutive mixtures of nonstationary signals which relies on the nonstationary nature of the sources to achieve separation. The algorithm is an on-line, LMS-like update based on minimizing the average squared cross-output-channel-correlations along with unity average energy output in each channel. We explain why, for nonstationary signals, such a criterion is sufficient to achieve source separation regardless of the signal statistics.

## 1. INTRODUCTION

The separation of multiple unknown sources from multi-sensor data has many applications, including the isolation of individual speech signals from a mixture of simultaneous speakers (as in video conferencing or the often-cited "cocktail party" environment), the elimination of cross-talk between horizontally and vertically polarized microwave communications transmissions, and the separation of multiple cellular telephone signals at a base station. In the past decade or so, a number of significant methods have been introduced for blind source separation, of which we review a few of the most popular here. One of the earliest and most effective methods (yet relatively unknown in some circles) is a constant-modulus-based method published in 1985 by Treichler and Larimore [1]. This method achieves simultaneous

separation and equalization by minimizing the deviation of the separated output magnitudes from a fixed gain. This method is very simple and convenient and works well even for non-constant-modulus signals with a sub-Gaussian kurtosis (which includes most communications signals).

Jutten and Herault introduced one of the most popular methods [2]. This method works well in many applications, particularly cross-talk situations in which a relatively modest amount of mixing occurs. For more challenging scenarios, the existence of multiple minima and misconvergence of the widely used Jutten-Herault algorithm has been examined in the literature [3]-[4]. Methods for non-Gaussian sources have also been developed, including [5] and others<sup>1</sup>. More recently, methods based on second-order statistics (and which can thus work even for Gaussian sources) have been introduced. A method by Belouchrani, *et al.* can separate stationary Gaussian sources with different autocorrelation statistics [6].

In many applications of blind source separation, the received signals are nonstationary. Nonstationarity may arise either from the source signals themselves (such as speech), or from channel impairments (such as fading in wireless communications channels). Most techniques for blind source separation assume stationarity of the signals and depend on reliable estimation of second-order or higher-order statistics. These methods may have difficulty when applied to nonstationary signals.

Several methods developed explicitly for nonstationary source separation have been published recently. Belouchrani and Amin have developed a time-frequency extension of the method in [7] for nonstationary sources, and Parra, *et al.* have developed another method based on frequency decomposition of several successive blocks of time [8]. While these methods appear effective, and

<sup>1</sup>It should be noted here that the CMA-based method by Treichler and Larimore also depends on the sub-Gaussianity of the sources.

This work was supported by the National Science Foundation, grant no. CCR-9979381.

the latter can also separate convolutive mixtures, they are block-based methods requiring somewhat sophisticated and expensive processing. Matsuoka, et al. present an on-line, adaptive extension of the Jutten-Herault method which, somewhat like the method we proposed in [9], attempts to minimize the average cross-correlation between separated channels while normalizing the output energy [10].

In various situations, convolutive (or dynamic) mixing occurs rather than instantaneous mixing. This complicates the BSS problem and requires a more sophisticated and computationally complex solution. Although the convolutive mixture problem is not as widely published as the instantaneous problem, methods for solving the problem are discussed in [11]–[12].

In this paper, we extend our work in [9] to convolutive mixtures to obtain a method for blind source separation of nonstationary, convolutively mixed signals which requires only nonstationarity and independence of the sources to achieve separation. An on-line, LMS-like algorithm is derived which achieves separation while normalizing the average energy of each output channel. This simple algorithm also offers tracking capability for time-varying convolutive mixtures. The optimization criterion is presented in the second section of this paper, the adaptive algorithm is derived in the third section, and simulations which illustrate its performance are presented in the fourth section. Some perspectives on the results are discussed in the final section.

## 2. A NONSTATIONARY CONVOLUTIVELY MIXED SOURCE SEPARATION CRITERION

The general source separation problem with convolutive mixtures can be described as

$$\mathbf{x}(n) = \sum_{m=-\infty}^n \mathbf{A}(n-m)\mathbf{s}(m), \quad (1)$$

where  $\mathbf{s}(n)$  is a vector of  $M$  zero-mean, statistically independent source processes at time-sample  $n$ ,  $\mathbf{x}(n)$  is a vector of  $N$  sensor measurements,  $N \geq M$ , and  $\mathbf{A}(n)$  is an  $M \times N$  mixing filter matrix. The goal of blind source separation is to determine an  $N \times M$  de-mixing matrix of filters  $\mathbf{B}(n)$  for  $n = 0 \dots L-1$ , which, when applied to the received sensor data as in

$$\mathbf{y}(n) = \sum_{m=0}^{L-1} \mathbf{B}(m)\mathbf{x}(n-m), \quad (2)$$

recovers (separates) the individual sources up to an unknown permutation and unknown channel gains, which

cannot be uniquely determined without additional information [10].

An important problem with convolutive mixtures is that even complete separation may not recover the exact original  $\mathbf{x}(n)$  source signals. Due to the blind nature of the problem and the memory introduced by the convolutive mixing, it may be impossible to obtain the true source signals, and instead filtered versions may result without further assumptions on the source signals. It is for this reason that convolutive-mixture-BSS-algorithm performance can be viewed, as in [11], by how well a system separates two sources without any regard to how the output signals compare to their unfiltered source versions. A way to quantify this separation performance is to see how well (statistically) uncorrelated the output signals are. In this paper though, our methods perform joint separation-equalization, and this should work well for a certain class of source signals. Our simulations compare the output signals to the original source signals and quantify the performance in terms of signal-to-interference ratio (SIR).

It has been observed in many papers on blind source separation that a necessary condition for the separation of zero-mean, statistically independent sources is that the cross-correlations of the output channels equal zero. However, this is not a sufficient condition, as is well known (see [9] for an example demonstrating this). For sources with fixed variances, an ambiguity exists as there are an infinite number of demixing matrices which obtain zero cross-channel correlation. For any arbitrary pair of variances, the classes of decorrelating matrices are *different* for different source variances, and only a true separating solution yields zero cross-channel correlation for *all* variance combinations. This is the key insight on which nonstationary blind source separation algorithms are based. In effect, these methods take multiple snapshots of the short-time cross-correlations at different times, and by minimizing all of these simultaneously, they exploit the changes in the relative channel variances to find a truly separating solution.

This paper uses the same basic insight, but proposes a new criterion for exploiting it which leads to a particularly simple and convenient algorithm. We propose to minimize the following criterion:

$$\min_{\substack{\mathbf{B}(n) \\ n=0 \dots L-1}} E \left[ \sum_{l=0}^{L-1} \sum_{i=1}^M \sum_{\substack{j=1 \\ j \neq i}}^M \hat{r}_{y_i y_j}^2(l) + \lambda \sum_{i=1}^M (\hat{r}_{y_i y_i}(0) - 1)^2 \right] \quad (3)$$

where at time  $n$

$$\hat{r}_{y_i y_j}(l; n) = \sum_k h(k) y_i(n-k-l) y_j(n-k) \quad (4)$$

and  $h(k)$  is a lowpass averaging filter for computing a short-term estimate of the cross-correlation of output channels  $y_i$  and  $y_j$  at time  $n$  and lag  $l$ . The first term in the criterion is to minimize the average squared magnitude of the short-term cross-correlations for the first  $L$  lags of the output signals (which, as discussed above and in [10], should only be achieved for nonstationary signals by a separating solution), while the second term demands that the output signals in each channel have unit energy on average. In a sense, the second criterion adds a signal normalization feature to the algorithm, but as was shown in [1], this CMA criterion has the ability to jointly separate and equalize sub-Gaussian signals. In instances where the source signals are sub-Gaussian (or one or more of them are), the added CMA criterion greatly aids in separating as well as equalizing in order to obtain closer estimates of the original  $\mathbf{x}(n)$  source signals.

### 3. ADAPTIVE ALGORITHM

There are many ways to construct a numerical algorithm based on the above criterion for blind nonstationary source separation, yielding different tradeoffs in terms of computational efficiency, convergence rate, block-based or adaptive forms, etc. However, in many applications, a simple, adaptive method which can track slow variations in the mixing parameters is desired. We derive here a stochastic gradient (LMS-like) algorithm which has these characteristics.

Many of the most successful adaptive algorithms are based on a stochastic gradient update using an instantaneous approximation to the expectation in the optimization criterion. For the optimization of the demixing matrices,  $\mathbf{B}(l)$ 's, a stochastic gradient update takes the form

$$\mathbf{B}_{n+1}(l) = \mathbf{B}_n(l) - \mu \nabla_n(l) \quad \text{for } l = 0 \dots L-1. \quad (5)$$

where

$$\nabla_n(l) = \left[ \frac{\partial}{\partial b_{pq}(l)} \left\{ \sum_{i=1}^M \sum_{\substack{j=1 \\ j \neq i}}^M \hat{r}_{y_i y_j}^2(l) + \lambda \sum_{i=1}^M (\hat{r}_{y_i y_i}(0) - 1)^2 \right\} \right] \quad (6)$$

where  $p$  and  $q$  are the row and column indices of the gradient matrix. Note the use of the instantaneous value at time  $n$  of the error function in (3) in the gradient computation. The  $(p, q)$ th element of the gradient matrix at lag  $l$  can easily be shown to be

$$\nabla_{pq,n}(l) = 2 \sum_{l=0}^{L-1} \left\{ \sum_{\substack{j=1 \\ j \neq p}}^M \hat{r}_{y_p y_j}(l) \hat{r}_{x_q y_j}(l-k) + \right.$$

$$\left. \sum_{\substack{i=1 \\ i \neq p}}^M \hat{r}_{y_i y_p}(l) \hat{r}_{y_i x_q}(-l-k) \right\} + 2\lambda(\hat{r}_{y_p y_p}(0) - 1)(\hat{r}_{y_p x_q}(k)) \quad (7)$$

We now derive efficient recursive updates for the short-term correlation estimates for a convenient form of the averaging filter. For computational efficiency, we select a first-order IIR averaging filter with impulse response

$$h(k) = \alpha^k u(k) \quad (8)$$

where  $u(k)$  is the unit step function and  $0 < \alpha < 1$ . With this form, the correlation statistics can easily be updated recursively according to

$$\tilde{r}_{y_i y_j}(l; n+1) = \alpha \tilde{r}_{y_i y_j}(l; n) + y_i(n-|l|)y_j(n), \quad (9)$$

and similarly

$$\tilde{r}_{y_i x_j}(l; n+1) = \alpha \tilde{r}_{y_i x_j}(l; n) + y_i(n-|l|)x_j(n) \quad (10)$$

for all lags  $l$  which are required for the algorithm. This completes the following simple recursive algorithm for nonstationary blind source separation.

1. Compute output according to (2).
2. Update short-time correlations using (9) and (10).
3. Compute separation filter gradient using (7).
4. Update separation filters as in (5).
5. Go back to step 1.

The complexity of the algorithm in the instantaneous mixture case was shown in [9] to be  $O(M^2 N)$ . Extension to the convolutive mixture case yields increased complexity by a factor of  $L^2$ , where  $L$  is of course a chosen parameter which can be used to trade off complexity and quality of separation.

### 4. SIMULATIONS

Several simulations have been performed to confirm the efficacy of the proposed method. For the following simulation with two sources and sensors, the mixing matrices are:

$$\mathbf{A} = \left\{ \begin{bmatrix} 1 & -.5 \\ .7 & 1.3 \end{bmatrix}, \begin{bmatrix} .35 & -.3 \\ -.2 & .6 \end{bmatrix}, \begin{bmatrix} -.2 & .2 \\ .15 & .3 \end{bmatrix} \right\} \quad (11)$$

where the first matrix represents zero lag, the second represents a lag of one, and the third represents a lag of two. The nonstationary sources, shown in Figure

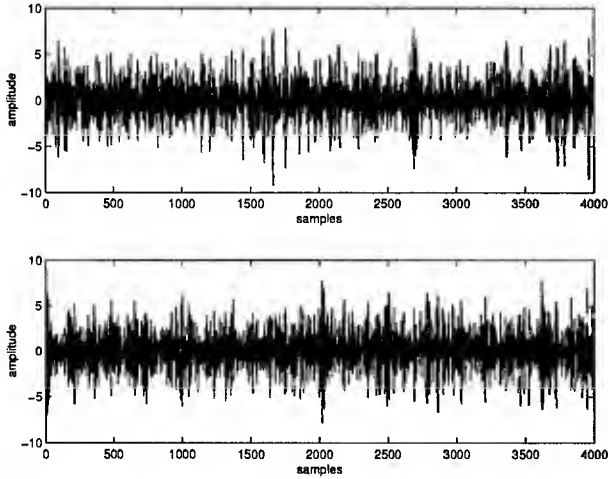


Figure 1: First 4000 samples of the nonstationary sources used in the simulation

1, are binary random signals multiplied by lowpass filtered Gaussian signals, and may be considered a crude approximation to communications signals undergoing fading. Three mixing scenarios are simulated by considering the cases of  $\mathbf{A}$  as above, only the first two matrices of  $\mathbf{A}$ , and only the first matrix of  $\mathbf{A}$  (ie. instantaneous mixture). These mixtures are tested against our source separation algorithm with  $L$  values ranging from 1 to 4, resulting in 12 different simulations.

Our BSS algorithm was tested in these 12 simulations and SIRs were computed for each of these cases, as well as for the case where no source separation is applied<sup>2</sup>. When our BSS algorithm is applied, output scaling is needed as BSS can only recover up to an unknown scale value. Since the scaling changes over time as the algorithm adapts, the signal was normalized by an approximate best-fit scale factor every 100 samples. A length-10,000 sample period was evaluated after sufficient convergence (using small values of  $\mu$ ) to obtain the resulting SIR values.

Table I shows the simulation results when only the first matrix in  $\mathbf{A}$  is used for mixing, which results in purely instantaneous mixing. The results show excellent performance for all cases of  $L$ , but one feature is that performance degrades slightly with increasing  $L$ . The reason for this is because only  $L = 1$  is needed to solve this problem, and by adding unneeded, adaptable coefficients, performance suffers slightly due to misadjustment in the stochastic gradient algorithm for the non-instantaneous coefficients.

Table II shows the simulation results for length-2 mixing (ie. only the first two matrices of  $\mathbf{A}$  are ap-

<sup>2</sup>In this case, the desired source signal is chosen according to which source is dominant in the mixture.

Table 1: Length-1 (Instantaneous) Mixture Results

BSS Type	SIR in dB	
	Source 1	Source 2
None	6.8954	6.5736
$L = 1$	36.1091	31.7012
$L = 2$	33.8167	32.2873
$L = 3$	32.0851	28.7450
$L = 4$	29.7797	28.9522

Table 2: Length-2 Mixing Results

BSS Type	SIR in dB	
	Source 1	Source 2
None	4.8529	4.6307
$L = 1$	13.3890	6.2799
$L = 2$	22.6628	11.3558
$L = 3$	27.6702	16.3054
$L = 4$	29.3789	21.2188

plied). The results clearly show a performance degradation compared to the instantaneous mixture results as the memory increases the difficulty of separation. It can be seen that the  $L = 1$  case does a fairly poor job of signal separation, and increasing  $L$  results in better SIR values as expected. Another observation is the imbalance of SIR performance between the two source signals. This is a function of the mixing filters.

Table III shows the simulation results for length-3 mixing using  $\mathbf{A}$  as in (11). The results show even further degradation than the length-2 mixture case as the increased mixing is more difficult to recover from. Again, the performance increases with the demixing filter length,  $L$ . Further gains could be obtained by using a larger  $L$ , but this comes at the expense of greater complexity of the system (proportional to  $L^2$ ) as well as much slower convergence.

Table 3: Length-3 Mixing Results

BSS Type	SIR in dB	
	Source 1	Source 2
None	4.4030	4.3170
$L = 1$	9.2718	5.7749
$L = 2$	11.0878	9.3144
$L = 3$	13.6907	12.4754
$L = 4$	15.8287	15.0418

## 5. CONCLUSIONS

Effective blind source separation can be achieved by exploiting nonstationarity of the sources. Furthermore, it is possible to separate convolutively mixed signals with the algorithm. This paper clearly shows performance gains can be made over an instantaneous mixture algorithm in the presence of convolutive mixtures.

Nonstationary blind source separation algorithms appear particularly relevant for practical applications because many sources of interest, such as speech or fading signals, exhibit nonstationarity but may not otherwise present features (such as non-Gaussian statistics or different auto-correlation structure) required by other methods.

In comparison with other nonstationary blind source separation algorithms, the method proposed here results in a simple on-line stochastic gradient algorithm requiring only multiplications and additions, which are efficiently implemented in signal processing hardware. It appears to exhibit the traditional characteristics of LMS-like algorithms including robustness and numerical stability, the ability to track slow variations in the environment, and relatively slow convergence.

The computational complexity of the algorithm is  $O(NM^2L^2)$ . That is, the cost is linear in the number of receivers, but quadratic in the number of sources and the demixing filter lengths. For many applications, these parameters are very small, and the algorithm is very efficient. For larger values of  $L$ , the computational cost may be the limiting factor in a tradeoff between performance and complexity.

## REFERENCES

- [1] J. R. Treichler and M. G. Larimore, "New processing techniques based on the constant modulus adaptive algorithm," *IEEE Transactions on Acoustics, Speech, and Signal Processing*, vol. 33, pp. 420-431, April 1985.
- [2] C. Jutten and J. Herault, "Blind separation of sources, part I: An adaptive algorithm based on neuromatic architecture," *Signal Processing*, vol. 24, pp. 1-10, July 1991.
- [3] P. Comon, C. Jutten, and J. Herault, "Blind separation of sources, part II: Problem statement," *Signal Processing*, vol. 24, pp. 11-20, July 1991.
- [4] Y. Srouchyari, "Blind separation of sources, part III: Stability analysis," *Signal Processing*, vol. 24, pp. 21-29, July 1991.
- [5] J.-F. Cardoso, "Iterative techniques for blind separation using only fourth-order cumulants," in *Signal Processing IV - Theories and Applications, Proceedings of EUSIPCO-92, Sixth European Signal Processing Conference*, vol. 2, pp. 739-742, 1992.
- [6] A. Belouchrani, K. A. Meraim, J.-F. Cardoso, and E. Moulines, "A blind source separation technique using second order statistics," *IEEE Transactions on Signal Processing*, vol. 45, pp. 434-444, February 1997.
- [7] A. Belouchrani and M. G. Amin, "Source separation based on the diagonalization of a combined set of spatial time-frequency distribution matrices," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP - 97*, (Germany), April 1997.
- [8] L. Parra, C. Spence, and B. de Vries, "Convolutive blind source separation based on multiple decorrelation," in *Proceedings of 1998 IEEE Workshop on Neural Networks for Signal Processing*, (Cambridge, UK), September 1998.
- [9] D. L. Jones, "A new method for blind source separation of nonstationary signals," in *Proceedings of 1999 IEEE International Conference on Acoustics, Speech, and Signal Processing, ICASSP - 99*, (Phoenix, AZ, USA), March 1999.
- [10] K. Matsuoka, M. Ohya, and M. Kawamoto, "A neural net for blind separation of nonstationary signals," *Neural Networks*, vol. 8, no. 3, pp. 411-419, 1995.
- [11] U. A. Lindgren and H. Broman, "Source separation using a criterion based on second-order statistics," *IEEE Transactions on Signal Processing*, vol. 46, pp. 1837-1850, July 1998.
- [12] H. L. N. Thi and C. Jutten, "Blind source separation for convolutive mixture," *Signal Processing*, vol. 45, pp. 209-229, 1995.

# A VERSATILE SPATIO-TEMPORAL CORRELATION FUNCTION FOR MOBILE FADING CHANNELS WITH NON-ISOTROPIC SCATTERING

A. Abdi, M. Kaveh

Dept. of Elec. and Comp. Eng., University of Minnesota  
Minneapolis, Minnesota 55455, USA

## ABSTRACT

For the analysis and design of adaptive antenna arrays in mobile fading channels, we need a model for the spatio-temporal correlation among the array elements. In this paper we propose a general spatio-temporal correlation function, where non-isotropic scattering is modeled by von Mises distribution, an empirically-verified model for non-uniformly distributed angle of arrival. The proposed correlation function has a closed form and is suitable for both mathematical analysis and numerical calculations. The utility of the new correlation function has been demonstrated by quantifying the effect of non-isotropic scattering on the performance of two applications of the antenna arrays for multiuser multichannel detection and single-user diversity reception. Comparison of the proposed correlation model with published data in the literature shows the flexibility of the model in fitting real data.

## 1. INTRODUCTION

In recent years the application of adaptive antenna arrays (smart antennas) for cellular systems has received much attention [1], since they can improve the coverage, quality, and capacity of such systems by combating interference, fading, and other undesired disturbances. An adaptive array can be defined as an adaptive spatio-temporal filter, which takes advantage of both time-domain and space-domain signal characteristics. Efficient joint use of time-domain and space-domain data demands a generalization of conventional communication theory and signal processing techniques to spatial and temporal communication theory [2] and space-time signal processing techniques [3]. Needless to say, new spatio-temporal channel models have to be developed as well. Since the second-order statistics of the channel characterize the basic structure of stochastic mobile channels, we need a spatio-temporal correlation function to study the basic impact of the random channel on the performance of space-time solutions, including the adaptive antenna arrays.

In this paper we present a flexible and versatile parametric correlation function for the mobile station (MS) (similar results can be obtained for the base station (BS) as well, as we see in Section 4). We do this by generalizing the spatio-temporal correlation function in [4], originally derived for an isotropic scattering scenario where the MS receives signals from all direction with equal probability, to the non-isotropic scattering case. Note that isotropic scattering at the MS corresponds to the uniform distribution for the angle of arrival (AOA) at the MS. However, empirical results have shown that due to the structure of the mobile channel, the MS is likely to receive signals only from particular directions (see [5] and references therein). In

other words, most often the MS experiences non-isotropic scattering, which results in a non-uniform distribution for the AOA at the MS. In [5] it has been shown that the application of von Mises distribution for the AOA at the MS yields an easy-to-use and closed-form expression for the temporal (or equivalently, spatial) correlation function. This correlation function has exhibited very good fit to measured data [5].

In the sequel we derive a new spatio-temporal correlation function where non-isotropic scattering is modeled by the von Mises distribution. To show the significant effect of non-isotropic scattering on the performance of smart antenna systems employing space-time data, we study the performance of an antenna array multiuser detector equipped with a channel estimator, operating in a Rayleigh fading channel. As a simpler example where only space data are employed, we also investigate the impact of non-isotropic scattering on a multi-element receiver working as a maximal ratio combiner (MRC) in a Rayleigh fading channel. In both examples we show how the proposed spatio-temporal correlation function helps us in quantifying the effect of the fading channel on the performance of antenna arrays, in the realistic scenario of non-isotropic scattering. The paper concludes with a comparison of the proposed correlation model with the published correlation data, collected by a BS-mounted array.

## 2. A NEW CORRELATION FUNCTION

Consider a linear uniformly-spaced antenna array shown in [4, Fig. 2], mounted on a MS. Let  $r_m(t)$  denote the complex envelope at the  $m$ th element from left. Then the normalized correlation function between the complex envelopes of the  $m$ th and the  $n$ th antenna elements, defined by  $\tilde{\phi}_{mn}(\tau) = E[r_m(t)r_n^*(t+\tau)]/E[|r_m(t)|^2]$ , can be derived from [4]:

$$\tilde{\phi}_{mn}(\tau) = E[\exp\{j2\pi f_d \tau \cos(\Theta - \alpha) + j(m-n)2\pi(d/\lambda) \cos \Theta\}], \quad (1)$$

where  $E$  denotes mathematical expectation,  $j = \sqrt{-1}$ ,  $f_d$  is the maximum Doppler frequency,  $\Theta$  stands for the AOA,  $\alpha$  represents the direction of the motion of the MS with respect to the horizontal axis counterclockwise,  $d$  is the spacing between any two adjacent antenna elements, and  $\lambda$  is the wavelength. Now we consider the von Mises probability density function (PDF) for the random variable  $\Theta$ :

$$p_\Theta(\theta) = \frac{\exp[\kappa \cos(\theta - \theta_p)]}{2\pi I_0(\kappa)}, \quad \theta \in [-\pi, \pi], \quad (2)$$

where  $I_0(\cdot)$  is the zero-order modified Bessel function,  $\theta_p \in [-\pi, \pi]$  accounts for the mean direction of AOA, and  $\kappa \geq 0$  controls the width of the AOA distribution [5]. For  $\kappa = 0$  (isotropic scattering) we have  $p_\Theta(\theta) = 1/(2\pi)$ , while for  $\kappa = \infty$

(extremely non-isotropic scattering) we obtain  $p_{\theta}(\theta) = \delta(\theta - \theta_p)$ , where  $\delta(\cdot)$  is the Dirac delta function. By calculating the expectation in (1) according to (2) we obtain:

$$I_0(\kappa) \tilde{\phi}_{mn}(\tau) = I_0 \left( \sqrt{\kappa^2 - x^2 - y^2 - 2xy \cos \alpha + j2\kappa [x \cos(\alpha - \theta_p) + y \cos \theta_p]} \right), \quad (3)$$

where  $x = 2\pi f_d \tau$  and  $y = 2\pi(m-n)d/\lambda$ . With  $\kappa = 0$ , (3) reduces to Lee's spatio-temporal correlation function  $J_0(\sqrt{x^2 + y^2 + 2xy \cos \alpha})$  in [4, Eqs. (42)-(43)] for isotropic scattering, where  $J_0(\cdot)$  is the zero-order Bessel function. For  $m = n = 1$  (single antenna), Lee's result further simplifies to Clarke's classic temporal correlation function  $J_0(x)$  [6, p. 40, Eq. (2.20)]. For a single antenna experiencing non-isotropic scattering and  $\alpha = 0$ , (3) reduces to the temporal correlation function  $I_0(\sqrt{\kappa^2 - x^2 + j2\kappa x \cos \theta_p})/I_0(\kappa)$  derived in [5, Eq. (2)] (this correlation function has shown very good fit to measured data [5]).

In comparison with the existing spatial correlation functions for antenna arrays [7], our proposed model in (3) has the main advantage that it includes both space and time dimensions in a single mathematically-tractable closed-form expression, flexible for fitting to array data, studying the performance of various array-based techniques [8] for different applications in fading channels with the realistic assumption of non-isotropic scattering, optimizing array configurations [9], etc..

### 3. TWO ARRAY APPLICATIONS

In this section we use the proposed model in (3) for two array-based applications. In the first one we need a spatio-temporal correlation function, while for the second one a spatial-only correlation function is needed. In array applications, the need for a spatio-temporal correlation function also appears in conjunction with such important fading characteristics as level crossing rate and average fade duration [4] [10], which due to space limitations we do not address here.

#### 3.1 Efficiency of Two Multiuser Multichannel Array Detectors

For code division multiple access (CDMA) signals, recently two array-based multiuser detection schemes with imperfect estimates of the fading channel were investigated in [11]: the decision-directed detector (with more complexity) which is optimum, and the decorrelating detector (with less complexity) which is suboptimum. In terms of the asymptotic efficiency, it has been proven that the decision-directed detector is superior. However, the decorrelating detector is simpler to implement. So, it is of interest to determine how much these two detectors are different in terms of asymptotic efficiency. Here, by a simple example [12, p. 107 and p. 117] we show that the answer strongly depends on the mode of scattering, which affects the correlation function of the complex envelope in the fading channel.

Assume that the MS has a two-element antenna ( $M = 2$ ), and there are two mobile users ( $K = 2$ ) according to the configuration shown in Figs. 1 and 2 ( $\theta_{p,1} = 0$ ,  $\theta_{p,2} = \pi$ ). In Fig. 1 we have  $\kappa_1 = \kappa_2 = 0$ , where the MS receives scattered plain

waves from all directions with equal probability, while in Fig. 2, where  $\kappa_1 = \kappa_2 = 10$ , the MS receives directional waves from two specific directions (the beamwidth in each direction is equal to  $\text{BW} = 2/\sqrt{\kappa} \approx 36^\circ$  [5]). Suppose the first user is the desired user, while the second one is the interfering user. The MS moves from left to right ( $\alpha = 0$ ) and the users travel at speeds such that the desired user has the maximum Doppler frequency  $f_{d,1} = 0.1$  Hz, while the interfering user has the maximum Doppler frequency  $f_{d,2} = 0.05$  Hz. Assume the correlation coefficient between the users' signature waveforms is  $\rho_{12} = 0.5$ , and the MS uses only the past two values ( $I = 2$ ) of matched filter outputs and bit decisions for fading estimation and bit detection in the presence of Rayleigh fading and zero-mean additive white Gaussian noise with variance  $\sigma^2$ . Suppose both users have (equal) unit power. Let us define the signal-to-noise ratio (SNR) as  $\gamma = 1/\sigma^2$ . For  $d = 0.3\lambda$  and  $\lambda$ , the asymptotic efficiency of the desired users,  $\eta_1$ , calculated using the equations given in [12], is plotted in Figs. 3 and 4 versus SNR, assuming  $\kappa_1 = \kappa_2 = 0$  and  $\kappa_1 = \kappa_2 = 10$ . According to both figures, as  $\kappa$  increases (more directional reception), the efficiency of both detectors increases significantly (which is good news). However, the difference between the detectors efficiencies increase as well, which implies that choosing the decorrelating detector, due to its lower complexity, introduces a significant loss in efficiency when we have non-isotropic scattering. Hence, we need to develop new suboptimum low-complexity detectors with efficiencies comparable with the optimum detector, in channels with directional reception.

#### 3.2 Average Bit Error Rate of a Single-User Multichannel Array Detector

Assume that in Figs. 1 and 2, we have user one only ( $K = 1$ ), and  $\theta_p = 0$ . Moreover, both the MS and the user are stationary ( $f_d = 0$ ). The user sends data using binary phase shift keying (BPSK) modulation scheme, and the MS is equipped with a two-branch ( $M = 2$ ) maximal ratio combiner (MRC). The average bit error rate (BER) in this case is given by [13, Eq. (12)]:

$$P_b(\gamma) = \frac{1}{4} \left\{ 2 - \frac{1}{\rho} \left[ (1+\rho) \sqrt{\frac{\gamma(1+\rho)}{1+\gamma(1+\rho)}} - (1-\rho) \sqrt{\frac{\gamma(1-\rho)}{1+\gamma(1-\rho)}} \right] \right\}, \quad (4)$$

where  $\rho = |\tilde{\phi}_{12}(0)|$ . In Figs. 5 and 6 we have plotted  $P_b(\gamma)$  versus  $\gamma$  for  $d = 0.3\lambda$  and  $\lambda$ , respectively. As we expect, the average BER increases as  $\kappa$  increases, because it results in more correlation between the branches. Of course, a larger  $d$  can reduce the amount of correlation between branches, resulting in smaller average BER (compare Figs. 5 and 6).

### 4. COMPARISON WITH DATA

Although the application of antenna arrays in both MS and BS is advantageous, in this section we focus on BS since the application of arrays at the BS is more common (practical constraints usually restrict the use of an array of antennas at a MS). For statistical characterization of narrow histograms of the AOA of waves impinging the BS [14] [15] (which gives rise to the non-uniform distribution of power versus the azimuth angle [16]), three different PDF's are used so far in the literature: cosine [17], Gaussian [18], and truncated uniform [19]. All these



PDF's are considered primarily for studying the effect of non-uniformly distributed AOA on the spatial correlation among the array elements at a BS. With appropriate choice of parameters, these three PDF's can resemble visually the narrow histograms of the AOA at the BS (although the truncated uniform PDF is less likely to do that because the empirical histograms are usually bell-shaped [14] [15] and decay to zero not as abruptly as a truncated uniform PDF). So, mathematical convenience seems to be the main concern in choosing a PDF for the AOA, among empirically-acceptable candidates. From this point of view, none of these three PDF's are able to provide a simple closed-form solution (in terms of known mathematical functions) for the correlation between the complex envelopes of the array elements (which is a basic quantity in array-related studies). For the Gaussian PDF only approximate results can be found [18] [20], and for the truncated uniform PDF, closed-form results can be derived only for inline and broadside cases [21] (the cosine PDF is less likely to yield a closed-form answer because of the special integral that has to be solved). On the other hand, as we see in the sequel, von Mises PDF yields a simple and compact expression, given in (5), which is basically the same as (3). This makes the von Mises PDF a very suitable model.

Comparison of the Gaussian PDF with the histograms of AOA data has shown reasonable agreement [15] [22]. This is a good empirical support for the von Mises PDF because for large  $\kappa$ , the PDF in (2) resembles a small-variance Gaussian PDF with mean  $\theta_p$  and standard deviation  $1/\sqrt{\kappa}$  [23, p. 60]. In fact, for any beamwidth (angle spread) smaller than  $40^\circ$  (which correspond to  $\kappa > 8.2$  according to the definition of beamwidth as  $\text{BW} = 2/\sqrt{\kappa}$  in [5]), the plots of Gaussian and von Mises PDF are indistinguishable (two typical standard deviations for the Gaussian PDF are  $15^\circ$  [22] and  $6^\circ$  [15], which correspond to  $\kappa = 14.6$  and  $\kappa = 91.2$ , respectively). However, recall that von Mises PDF is able to provide a general and closed-form solution for the space-time correlation between the complex envelopes of the array elements, while Gaussian PDF cannot.

Using exactly the same notation as [17], it is straightforward to show that for the linear uniformly-spaced antenna array at the BS in [17, Fig. 6] we have:

$$I_0(\kappa) \tilde{\phi}_{mn}(\tau) = I_0 \left( \sqrt{\kappa^2 - x^2 - y^2 + 2xy \cos \gamma + j2\kappa[x \cos(\gamma - \alpha) - y \cos \alpha]} \right), \quad (5)$$

provided that AOA has a von Mises PDF with the mean direction  $\alpha \in [-\pi, \pi)$  and the width control parameter  $\kappa \geq 0$ . All of the parameters in (5) are the same as (3), except for  $\gamma$  in (5) which represents the direction of the motion of the MS with respect to the horizontal axis counterclockwise, in place of  $\alpha$  in (3) (the  $\gamma$  here should not be confused with the SNR symbol  $\gamma$ , used in Section 3). The two sign changes in (5), in comparison with (3), come from different ways of numbering the array elements: in [4, Fig. 2], the elements are numbered from left to right, while elements numbering in [17, Fig. 6] is from right to left.

Now we compare our correlation model with the data published in [17], where the data are spatial cross-correlations between the square of the envelopes of a two element array, mounted on a BS. We do this by considering two models for the AOA PDF at the BS: the simple model with

$p_\theta(b) = \exp\{\kappa \cos(b - \alpha)\} / 2\pi I_0(\kappa)$ , and the composite model with  $p_\theta(b) = \zeta \exp\{\kappa \cos(b - \alpha)\} / 2\pi I_0(\kappa) + (1 - \zeta) / 2\pi$ , where  $0 \leq \zeta \leq 1$  indicates the amount of directional reception. The composite PDF reduces to the von Mises PDF for  $\zeta = 1$ , and simplifies to the uniform PDF for  $\zeta = 0$ . Consequently, the associated spatial correlation functions for a two element array at a BS can be written as:

$$\begin{aligned} \tilde{\phi}_{12}(0) &= I_0 \left( \sqrt{\kappa^2 - 4\pi^2(d/\lambda)^2 + j4\pi\kappa(d/\lambda)\cos\alpha} \right) / I_0(\kappa), \quad (6) \\ \tilde{\phi}_{12}(0) &= \zeta I_0 \left( \sqrt{\kappa^2 - 4\pi^2(d/\lambda)^2 + j4\pi\kappa(d/\lambda)\cos\alpha} \right) / I_0(\kappa) \\ &\quad + (1 - \zeta) J_0(2\pi d/\lambda). \quad (7) \end{aligned}$$

Figs. 7-8 show Lee's correlation data, plotted together with  $|\phi_{12}(0)|^2$  calculated according to (6) and (7) for both models. For a given  $\alpha$  (known a priori for each data set), the unknown  $\kappa$  for the simple model and the unknown pair  $(\kappa, \zeta)$  for the composite model are estimated by the nonlinear least squares method (implemented via a systematic numerical search technique). Based on these figures (and many others not shown due to space limitations), the von Mises PDF is able to account for the variations of the correlation versus antenna spacing with reasonable accuracy (compare our correlation plots with those drawn in [17] assuming the cosine PDF and [21] using the truncated uniform PDF, both for the same data sets. Interestingly, the correlation plots in [17] can also be considered as curves obtained based on a Gaussian PDF, because for small BW, the cosine PDF can be approximated by a Gaussian PDF [21]). Note that in Fig. 7 both models are similar ( $\zeta = 0.98$ ), while in Fig. 8 the composite model shows a much better fit ( $\zeta = 0.74$ ). In general the composite model was able to improve the fits obtained by the simple model, which is not surprising because it has the additional parameter  $\zeta$ . This is in agreement with the noise-like signal introduced in [17].

## 5. CONCLUSION

Space-time processing using antenna arrays over wireless mobile fading channels offer several advantages in cellular systems, such as mitigating fading, intersymbol interference, cochannel interference, etc.. Efficient joint use of both space and time dimensions demands for spatio-temporal channel models. As a basic channel model, we need a two dimensional spatio-temporal correlation function among the random signals sensed by the array elements, to characterize the second order dependence structure of the random channel in both space and time. In this paper we have proposed a flexible spatio-temporal correlation function for propagation scenarios with non-isotropic scattering (signal reception from specific directions). The non-uniform distribution for the angle of arrival, which characterizes the non-isotropic scattering, is modeled by von Mises PDF which has previously shown to be successful in describing the measured data. The proposed spatio-temporal correlation function is general enough to include important special cases such as Lee's spatio-temporal correlation function and Clarke's temporal correlation function, both derived for isotropic scattering. Moreover, its compact mathematical form facilitates analytical manipulations of array-based techniques and results in terms of closed-form expressions for such important fading parameters as



spectral moments (successive derivatives of the correlation function). Based on two case studies (multiuser detection and diversity reception) and using the new spatio-temporal correlation function, we have shown that non-isotropic scattering (typical of many mobile channel scenarios) has a significant impact on the performance of array processors, and should be taken into account in the analysis and design of adaptive antenna arrays for mobile fading channels.

Theoretically, the new correlation function is applicable to both MS and BS. However, since practical restrictions limit the use of multiple antennas at a MS, the proposed correlation function seems to be of much more use in a BS. Therefore, the empirical justification of the new correlation function is demonstrated by comparison with published data collected at a BS.

## 6. ACKNOWLEDGEMENT

This work has been supported in part by the National Science Foundation, under the Wireless Initiative Program, Grant #9979443. The authors appreciate the input provided by Dr. T. A. Brown at Motorola regarding the multiuser multichannel detector examples.

## 7. REFERENCES

- [1] J. H. Winters, "Smart antennas for wireless systems," *IEEE Pers. Commun. Mag.*, vol. 5, no. 1, pp. 23-27, 1998.
- [2] R. Kohno, "Spatial and temporal communication theory using adaptive antenna array," *IEEE Pers. Commun. Mag.*, vol. 5, no. 1, pp. 28-35, 1998.
- [3] A. J. Paulraj and C. B. Papadias, "Space-time processing techniques for wireless communications," *IEEE Signal Processing Mag.*, vol. 14, no. 6, pp. 49-83, 1997.
- [4] W. C. Y. Lee, "Level crossing rates of an equal-gain predetection diversity combiner," *IEEE Trans. Commun. Technol.*, vol. 18, pp. 417-426, 1970.
- [5] A. Abdi, H. Allen Barger, and M. Kaveh, "A parametric model for the distribution of the angle of arrival and the associated correlation function and power spectrum at the mobile station," submitted to *IEEE Trans. Vehic. Technol.*, Sep. 1999.
- [6] G. L. Stuber, *Principles of Mobile Communication*. Boston, MA: Kluwer, 1996.
- [7] R. B. Ertel, P. Cardieri, K. W. Sowerby, T. S. Rappaport, and J. H. Reed, "Overview of spatial channel models for antenna array communication systems," *IEEE Pers. Commun. Mag.*, vol. 5, no. 1, pp. 10-22, 1998.
- [8] L. C. Godara, "Applications of antenna arrays to mobile communications, Part I: Performance improvement, feasibility, and system considerations, Part II: Beam-forming and direction-of-arrival considerations," *Proc. IEEE*, vol. 85, pp. 1031-1060 and pp. 1195-1245, 1997.
- [9] W. C. Y. Lee, "A study of the antenna array configuration of an  $M$ -branch diversity combining mobile radio receiver," *IEEE Trans. Vehic. Technol.*, vol. 20, pp. 93-104, 1971.
- [10] F. Adachi, M. T. Feeney, and J. D. Parsons, "Effects of correlated fading on level crossing rates and average fade durations with predetection diversity reception," *IEE Proc. F, Commun., Radar, Signal Processing*, vol. 135, pp. 11-17, 1988.
- [11] T. A. Brown and M. Kaveh, "Multiuser detection with antenna arrays in the presence of multipath fading," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, Atlanta, GA, 1996, pp. 2662-2665.
- [12] T. A. Brown, "The use of antenna arrays in the detection of code division multiple access signals," Ph.D. Thesis, Dept. of Elec. Eng., University of Minnesota, Minneapolis, MN, June 1995.
- [13] S. T. Kim, J. H. Yoo, and H. K. Park, "A spatially and temporally correlated fading model for array antenna applications," *IEEE Trans. Vehic. Technol.*, vol. 48, pp. 1899-1905, 1999.
- [14] A. Klein and W. Mohr, "A statistical wideband mobile radio channel model including the directions-of-arrival," in *Proc. IEEE Int. Symp. Spread Spectrum Techniques Applications*, Mainz, Germany, 1996, pp. 102-106.
- [15] K. I. Pedersen, P. E. Mogensen, and B. H. Fleury, "A stochastic model of the temporal and azimuthal dispersion seen at the base station in outdoor propagation environments," *IEEE Trans. Vehic. Technol.*, vol. 49, pp. 437-447, 2000.
- [16] P. Pajusco, "Experimental characterization of D.O.A at the base station in rural and urban area," in *Proc. IEEE Vehic. Technol. Conf.*, Ottawa, ONT, Canada, 1998, pp. 993-997.
- [17] W. C. Y. Lee, "Effects on correlation between two mobile radio base-station antennas," *IEEE Trans. Commun.*, vol. 21, pp. 1214-1224, 1973.
- [18] F. Adachi, M. T. Feeney, A. G. Williamson, and J. D. Parsons, "Crosscorrelation between the envelopes of 900 MHz signals received at a mobile radio base station site," *IEE Proc. F, Commun., Radar, Signal Processing*, vol. 133, pp. 506-512, 1986.
- [19] J. Salz and J. H. Winters, "Effect of fading correlation on adaptive arrays in digital mobile radio," *IEEE Trans. Vehic. Technol.*, vol. 43, pp. 1049-1057, 1994.
- [20] T. Trump and B. Ottersten, "Estimation of nominal direction of arrival and angular spread using an array of sensors," *Signal Processing*, vol. 50, pp. 57-69, 1996.
- [21] M. Kalkan and R. H. Clarke, "Prediction of the space-frequency correlation function for base station diversity reception," *IEEE Trans. Vehic. Technol.*, vol. 46, pp. 176-184, 1997.
- [22] U. Martin, "Spatio-temporal radio channel characteristics in urban macrocells," *IEE Proc. Radar, Sonar, Navig.*, vol. 145, pp. 42-49, 1998.
- [23] K. V. Mardia, *Statistics of Directional Data*. London: Academic, 1972.

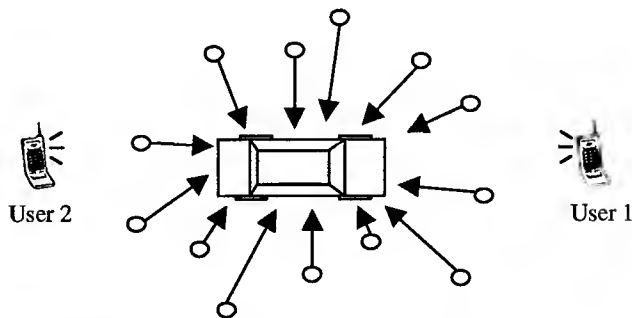


Figure 1. Isotropic scattering in an open area (circles are scatterers).

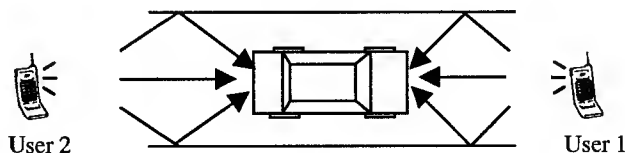


Figure 2. Non-isotropic scattering in a narrow street.

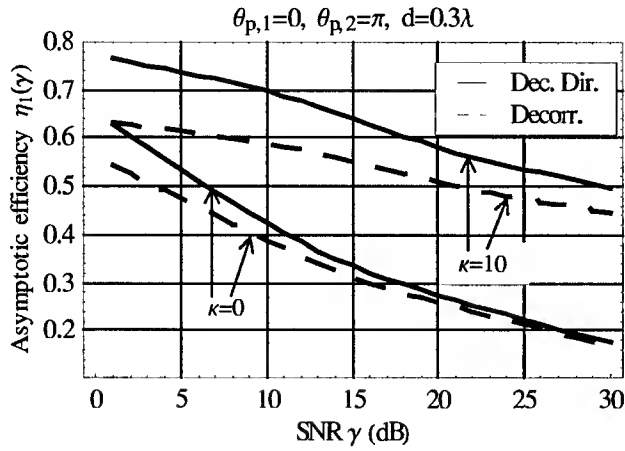


Figure 3. Asymptotic efficiency of two multiuser array detectors.

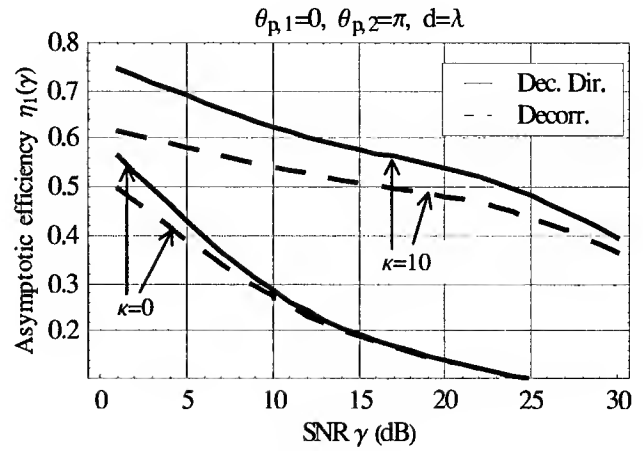


Figure 4. Asymptotic efficiency of two multiuser array detectors.

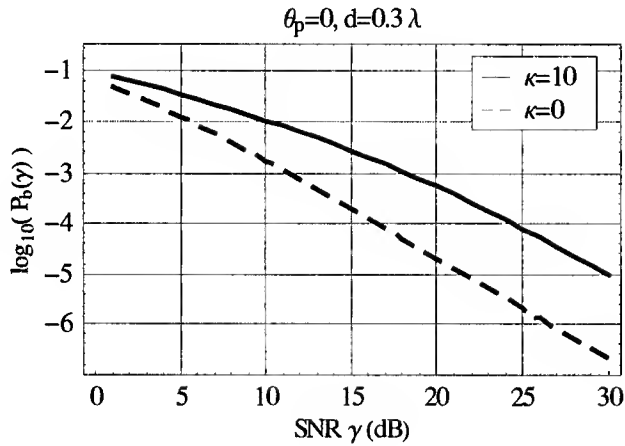


Figure 5. Bit error rate of BPSK with two-branch MRC.

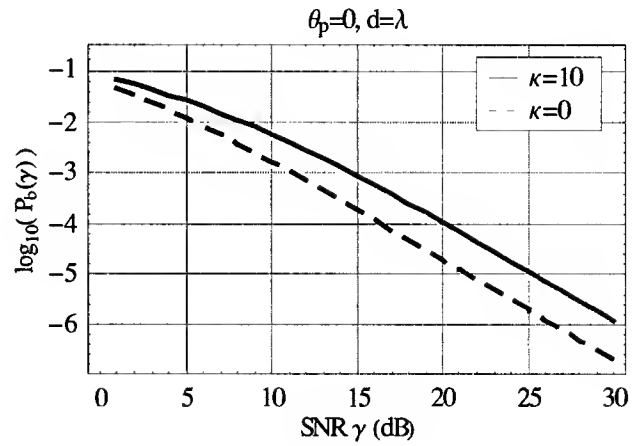


Figure 6. Bit error rate of BPSK with two-branch MRC.

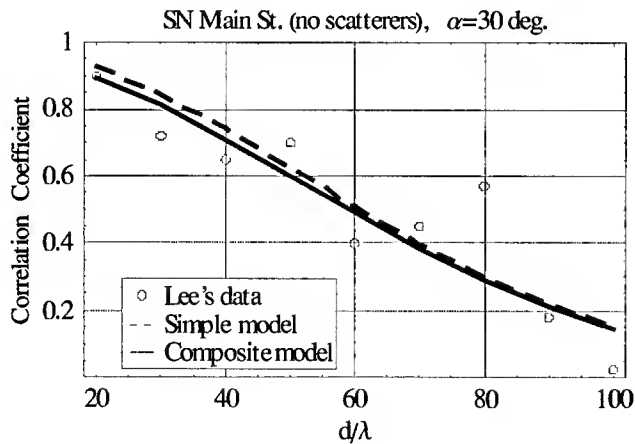


Figure 7. Correlation coefficient versus antennas spacing  
Simple: BW = 0.5°, Composite: BW = 0.5°,  $\zeta = 0.98$

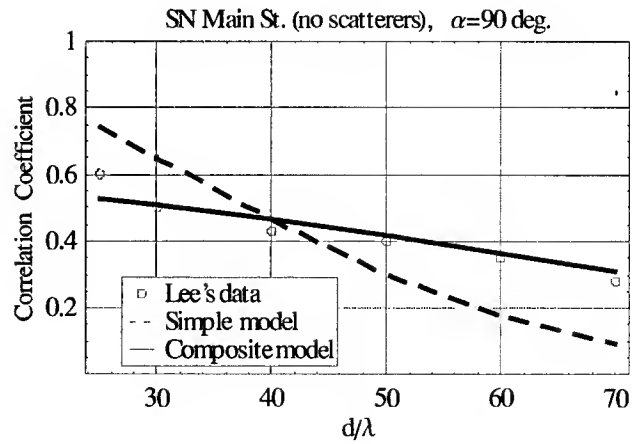


Figure 8. Correlation coefficient versus antennas spacing  
Simple: BW = 0.4°, Composite: BW = 0.2°,  $\zeta = 0.74$

# A BATCH SUBSPACE ICA ALGORITHM.

Ali MANSOUR and Noboru OHNISHI

Bio-Mimetic Control Research Center (RIKEN),  
2271-130, Anagahora, Shimoshidami, Moriyama-ku, Nagoya 463 (JAPAN)  
email:mansour@nagoya.riken.go.jp and ohnishi@ohnishi.nuie.nagoya-u.ac.jp  
http://www.bmc.riken.go.jp

## ABSTRACT

For the blind separation of sources (BSS) problem (or the independent component analysis (ICA)), it has been shown in many situations, that the adaptive subspace algorithms are very slow and need an important computation efforts. In a previous publication, we proposed a modified subspace algorithm for stationary signals. But that algorithm was limited to stationary signals and its convergence was not fast enough.

Here, we propose a batch subspace algorithm. The experimental study proves that this algorithm is very fast but its performance are not enough to completely achieve the separation of the independent component of the signals. In the other hand, this algorithm can be used as a pre-processing algorithm to initialize other adaptive subspace algorithms. **Keywords:** blind separation of sources, ICA, subspace methods, Lagrange method, Cholesky decomposition.

## 1. INTRODUCTION

The blind separation of sources (BSS) problem [1] (or the Independent Component Analysis "ICA" problem [2]) is a recent and important problem in signal processing. According to this problem, one should estimate, using the output signals of an unknown channel (i.e. the observed signals or the mixing signals), the unknown input signals of that channel (i.e. sources). The sources are assumed to be statistically independent from each other.

At first the BSS was proposed in a biological context [3]. Actually, one can find this problem in many different situations: speech enhancement [4], separation of seismic signals [5], sources separation method applied to nuclear reactor monitoring [6], airport surveillance [7], noise removal from biomedical signals [8], etc.

Since 1985, many researchers have been interested in BSS [9, 10, 11, 12]. Most of the algorithms deal with a linear channel model: The instantaneous mixtures (i.e. memoryless channel) or the convolutive mixtures (i.e. the channel effect can be considered as a linear filter). The criteria of those algorithms were generally based on high order statistics [13, 14, 15]. Recently, by using only second order statistics, some subspace methods have been explored to separate blindly the sources in the case of convolutive

mixtures [16, 17].

In previous works, we proposed two subspace approaches using LMS [18, 17] or a conjugate gradient algorithm [19] to minimize subspace criteria. Those criteria were been derived from the generalization of the method proposed by Gesbert *et al.* [20] for blind identification<sup>1</sup>. To improve the convergence speed of our algorithms, we proposed a modified subspace algorithm for stationary signals [21]. But that algorithm was limited to stationary signals and its convergence was not fast enough. Here, we propose a new subspace algorithm, which improves the performance of our previous methods.

## 2. MODEL, ASSUMPTIONS & CRITERION

Let  $Y(n)$  denotes the  $q \times 1$  mixing vector obtained from  $p$  unknown and statistically independent sources  $S(n)$  and let the  $q \times p$  polynomial matrix  $\mathcal{H}(z) = (h_{ij}(z))$  denotes the channel effect (see fig. 1). In this paper, we assume that the filters  $h_{ij}(z)$  are causal and finite impulse response (FIR) filters. Let us denote by  $M$  the highest degree<sup>2</sup> of the filters  $h_{ij}(z)$ . In this case,  $Y(n)$  can be written as:

$$Y(n) = \sum_{i=0}^M \mathbf{H}(i)S(n-i), \quad (1)$$

where  $S(n-i)$  is the  $p \times 1$  source vector at the time  $(n-i)$  and  $\mathbf{H}(i)$  is the real  $q \times p$  matrix corresponding to the filter matrix  $\mathcal{H}(z)$  at time  $i$ .

Let  $Y_N(n)$  (resp.  $S_{M+N}(n)$ ) denotes the  $q(N+1) \times 1$  (resp.  $(M+N+1)p \times 1$ ) vector given by:

$$Y_N(n) = \begin{pmatrix} Y(n) \\ \vdots \\ Y(n-N) \end{pmatrix},$$

$$S_{M+N}(n) = \begin{pmatrix} S(n) \\ \vdots \\ S(n-M-N) \end{pmatrix}.$$

<sup>1</sup>In the identification problem, the authors generally assume that they have one source and that the source is an iid signal.

<sup>2</sup> $M$  is called the degree of the filter matrix  $\mathcal{H}(z)$ .

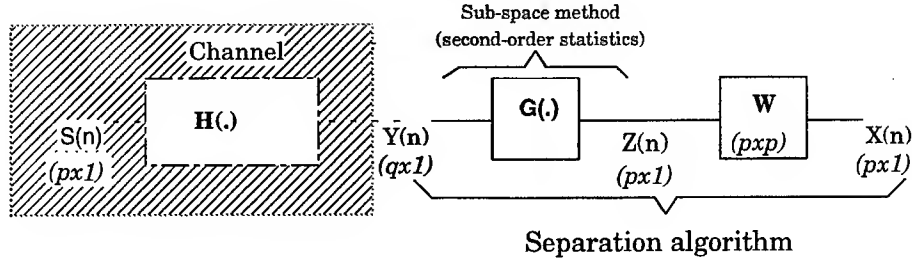


Figure 1: General Structure.

By using  $N > q$  observations of the mixture vector, we can formulate the model (1) in another form:

$$Y_N(n) = T_N(H)S_{M+N}(n), \quad (2)$$

where  $T_N(H)$  is the Sylvester matrix corresponding to  $H(z)$ . The  $q(N+1) \times p(M+N+1)$  matrix  $T_N(H)$  is given by [22] as:

$$\begin{bmatrix} H(0) & H(1) & \dots & H(M) & 0 & \dots & 0 \\ 0 & H(0) & \dots & H(M-1) & H(M) & 0 & \dots \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & H(0) & H(1) & \dots & H(M) \end{bmatrix}$$

It was proved in [23] that the rank of Sylvester matrix  $T_N(H) = p(N+1) + \sum_{i=1}^p M_i$ , where  $M_i$  is the degree of the  $i$ th column<sup>3</sup> of  $H(z)$ . Now, it is easy to prove that the Sylvester matrix has a full rank and it is left invertible if each column of the polynomial matrix  $H(z)$  has the same degree and  $N > Mp$  (see [24] for more details). From equation (2), one can conclude that the separation of the sources can be achieved by estimating a  $(M+N+1)p \times q(N+1)$  left inverse matrix  $G$  of the Sylvester matrix. To estimate  $G$ , one can use criterion proposed in [17] obtained from the generalization of the criterion in [20]:

$$\min C(G) = E \|(I \ 0)GY_N(n) - (0 \ I)GY_N(n+1)\|^2, \quad (3)$$

here  $E$  stands for the expectation,  $I$  is the identity matrix and  $0$  is a zero matrix of appropriate dimensions. It has been shown in [17] that the above minimization lead us to a matrix  $G^*$  such:

$$\text{Perf} = G^*T_N(H) = \text{diag}(M, \dots, M), \quad (4)$$

where  $M$  is any  $p \times p$  matrix. Using the last equation, it becomes clear that the separation is reduced to the separation of an instantaneous mixture with a mixing matrix  $M$ . In other words, this algorithm can be decomposed into two steps: First step, by using only second-order statistics, we reduce the convolutive mixture problem to an instantaneous mixture (deconvolution step); then in the second step, we must only separate sources consisting of a simple instantaneous mixture (typically, most of the instantaneous mixture algorithms are based on fourth-order statistics).

<sup>3</sup>The degree of a column is defined as the highest degree of the filters in this column.

Finally, to avoid the spurious solutions (i.e. a singular matrix  $M$ ), one must minimize that criterion subject to a constraint [17]:

$$\text{Subject to } G_0 R_N(n) G_0^T = I, \quad (5)$$

here  $R_N(n) = E Y_N(n) Y_N^T(n)$ , and the  $p \times q(N+1)$  matrix  $G_0$  stands for the first bloc line of  $G = (G_0^T \dots G_{(M+N)}^T)^T$ . The minimization using a LMS algorithm of the above criterion with respect to a constraint was discuss in our previous work [17]. In addition, the minimization of a modified version of the above criterion was done using a conjugate gradient algorithm [19].

### 3. ALGORITHM

From the previous section, it is clear that the minimization of the criterion (3) should be done subject to a  $p^2$  constraints<sup>4</sup>. Let  $\text{const}$  denotes the constraint vector (i.e.  $\text{const} = \text{Vec}(G_0 R_N(n) G_0^T - I)$ , here  $\text{Vec}$  is the operator that corresponds to a  $p \times q$  matrix a  $pq$  vector). The minimization of the criterion (3) subject to the constraints (5) can be formulated using the Lagrange method as:

$$\mathcal{L}(G, \lambda) = C(G) - \lambda \text{const} \quad (6)$$

here  $\lambda$  is a line vector, stands for the Lagrange parameters. The minimization of the above equation with respect to  $\lambda$  leads us to the constraint equation (5). Using the derivative  $\partial C(G)/\partial G$  given in [17], the equation (5) and (6), one can write:

$$\begin{aligned} \frac{\partial \mathcal{L}(G, \lambda)}{\partial G} = & \begin{pmatrix} I_p & 0 & 0 \\ 0 & 2I_{(M+N-1)p} & 0 \\ 0 & 0 & I_p \end{pmatrix} G R_N(n) \\ & - \begin{pmatrix} 0 & I_{(M+N)p} \\ 0 & 0 \end{pmatrix} G R_N^T(n+1) \\ & - \begin{pmatrix} 0 & 0 \\ I_{(M+N)p} & 0 \end{pmatrix} G R_N(n+1) - \begin{pmatrix} 2\Gamma & G_0 R_N(n) \\ 0 & 0 \end{pmatrix}, \end{aligned}$$

where  $R_N(n+1) = E Y_N(n) Y_N^T(n+1)$  and  $I_l$  is the  $l \times l$  identity matrix. By canceling the above equation and after some algebraic operations, one can find that the bloc lines

<sup>4</sup>Using the symmetrical form of the equation (5), one can decrease the constraint number to  $p(p+1)/2$ .

of the optimal  $\mathbf{G}^*$  should satisfy:

$$\mathbf{G}_0 \mathbf{R}_N(n) \mathbf{G}_0^T = \mathbf{I}, \quad (7)$$

$$2\mathbf{G}_i \mathbf{R}_N(n) = \mathbf{G}_{(i+1)} \mathbf{R}_N(n+1) + \mathbf{G}_{(i-1)} \mathbf{R}_N(n+1), \quad (8)$$

$$\mathbf{G}_{(M+N)} \mathbf{R}_N = \mathbf{G}_{(M+N-1)} \mathbf{R}_N(n+1), \quad (9)$$

here  $1 \leq i \leq M+N-1$ . Let  $\mathbf{A} = \mathbf{R}_N^T(n+1) \mathbf{R}_N^{-1}(n)$  and  $\mathbf{B} = \mathbf{R}_N^T(n+1) \mathbf{R}_N^{-1}(n)$ , we should mention that  $\mathbf{A}$  and  $\mathbf{B}$  exist if and only if (iff)  $\mathbf{R}_N(n)$  is full rank<sup>5</sup>. Finally, using some algebraic operations, we can prove that the previous matrix equation system can be solved by a recursion formula:

$$\mathbf{G}_{(M+N-i-1)} = \mathbf{G}_{(M+N-i-2)} \mathbf{D}_i \quad (10)$$

here  $0 \leq i \leq M+N-1$  and the  $\mathbf{G}_0$  can be obtained from the first equation (7), using a simple Cholesky decomposition. In addition, the matrices  $\mathbf{D}_i$  can also be obtained by:

$$\mathbf{D}_{(i+1)} = \mathbf{B}(2\mathbf{I} - \mathbf{D}_i \mathbf{A})^{-1} \quad (11)$$

here  $0 \leq i \leq M+N-1$  and  $\mathbf{D}_0 = \mathbf{B}$ . Even if relationships (10) and (11) looks complicated, but the time needed to obtain the matrix  $\mathbf{G}$  still very comparable<sup>6</sup> to the time needed for the convergence of LMS version [17] or even the Conjugate Gradient version [21, 19].

#### 4. EXPERIMENTAL RESULTS

The experiments discussed here are conducted using two sources ( $p = 2$ ) with uniform probability density function (pdf) and four sensors ( $q = 4$ ), and the degree of  $\mathcal{H}(z)$  is chosen as ( $M = 4$ ).

To show the performances of the subspace criterion, the matrix  $\text{Perf} = \mathbf{G}^* \mathbf{T}_N(\mathbf{H})$  is plotted. In the other hand, we know that the deconvolution is achieved iff the matrix  $\text{Perf}$  is a bloc diagonal matrix as shown in equation (4). Figure 2 shows the performances of the batch subspace algorithm discussed in this paper. It is clear from that figure 2 that the first step of the algorithm (the deconvolution) was not satisfactory achieved ( $\text{Perf}$  is not a bloc diagonal as in equation (4)). This problem was obtained because the criterion (3) is a flat function around its minima (see figure (2)).

Figure 3 shows us the performance results and the criterion convergence of the LMS algorithm (first column), and the performance results and the criterion convergence of

the same LMS algorithm but the matrix  $\mathbf{G}$  is initialized using the result of the batch algorithm (second column). We should mention that the time needed to obtain the minima by the initialized version was almost half the time needed by the non initialized version. Figures 3 (c) and (d) show the criterion convergence (the stop condition was the limit of the sample number, i.e. 10000). The experimental studies show that the Conjugate Gradient version of the subspace algorithm can converge faster and lead us to better performances if that algorithm has been initialized using the batch proposed algorithm (these results will be omitted in this short paper).

The second step of the algorithm consists on the separation of a residual instantaneous mixture (corresponding to  $\mathbf{M}$ , see equation (4)). This separation can be processed using any source separation algorithm applicable to instantaneous mixtures. Here, we chose the minimization of a cross-cumulant criterion using Levenberg-Marquardt method [25]. Figure (4) shows us the different signals (see figure (1)). It is clear that the sources  $X$  and the estimated signals  $S$  are independent signals and the vector  $Z$ , output of the subspace criterion, corresponds to an instantaneous mixture, and the observed vector  $Y$  corresponds to a convolutive mixture (see [26, 27]).

Finally, the estimation of the second and the high order statistics was done according to the method described in [28].

#### 5. CONCLUSION

In this paper, we propose a batch algorithm for source separation in convolutive mixtures based on a subspace approach. This new algorithm requires, as same as the other subspace methods, that the number of sensors is larger than the number of sources. In addition, it allows the separation of convolutive mixtures of independent sources using mainly second-order statistics: A simple instantaneous mixture, the separation of which generally needs high-order statistics, should be conducted to achieve the separation.

The experimental study shows that the the present algorithm can be used for initialized an adaptive subspace algorithm. The initialized algorithms need less time to converge. These results were discussed in the case of two subspace algorithms which are based on LMS or on a conjugate gradient method. Finally, the subspace LMS criterion and the Conjugate gradient criterion will become more stable and faster if they are initialized using the present algorithm.

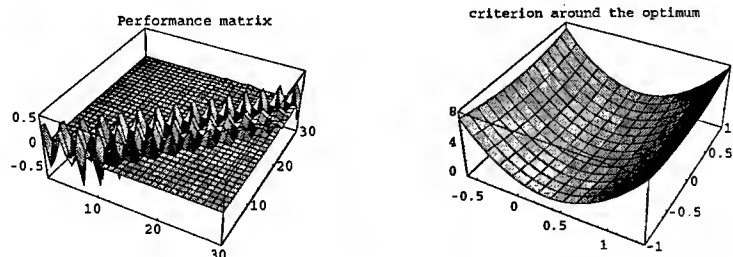
#### REFERENCES

- [1] C. Jutten and J. Héroult, "Blind separation of sources, Part I: An adaptive algorithm based on a neuromimetic architecture," *Signal Processing*, vol. 24, no. 1, pp. 1-10, 1991.
- [2] P. Comon, "Independent component analysis, a new concept?," *Signal Processing*, vol. 36, no. 3, pp. 287-314, April 1994.

<sup>5</sup>It is easy to prove that  $\mathbf{R}_N(n)$  is full rank iff one add some additive independent noise to the observed signals, because one of the subspace assumption  $q > p$ . In the other hand and by using the criterion (3), one can prove the existence of some spurious minima, if the model have some additive noise (the demonstration will be omitted here because the limit of the sheet number). However, the experimental study shows that one still obtain good results for a 20 dB ratio of signal to noise (RSN). In our simulation, we added a Gaussian noise with  $\text{RSN} \geq 20\text{dB}$ .

<sup>6</sup>Indeed, using C code program and an ultra 30 creator sun station, it needs few minutes (less than 5) to obtained the matrix  $\mathbf{G}$ . But the convergence of the conjugate gradient needs from 40 to 100 minutes and the LMS algorithm needs few hours to converge.

- [3] B. Ans, J. C. Gihodes, and J. Hérault, "Simulation de réseaux neuronaux (sirene). II. hypothèse de décodage du message de mouvement porté par les afférences fusorales IA et II par un mécanisme de plasticité synaptique," *C. R. Acad. Sci. Paris*, vol. série III, pp. 419–422, 1983.
- [4] L. Nguyen Thi and C. Jutten, "Blind sources separation for convolutive mixtures," *Signal Processing*, vol. 45, no. 2, pp. 209–229, 1995.
- [5] N. Thirion, J. MARS, and J. L. BOELLE, "Separation of seismic signals: A new concept based on a blind algorithm," in *Signal Processing VIII, Theories and Applications*, Triest, Italy, September 1996, pp. 85–88, Elsevier.
- [6] G. D'urso and L. Cai, "Sources separation method applied to reactor monitoring," in *Proc. Workshop Athos working group*, Girona, Spain, June 1995.
- [7] E. Chaumette, P. Common, and D. Muller, "Application of ica to airport surveillance," in *HOS 93*, South Lake Tahoe-California, 7-9 June 1993, pp. 210–214.
- [8] A. Kardec Barros, A. Mansour, and N. Ohnishi, "Removing artifacts from ecg signals using independent components analysis," *NeuroComputing*, vol. 22, pp. 173–186, 1999.
- [9] J. F. Cardoso and P. Comon, "Tensor-based independent component analysis," in *Signal Processing V, Theories and Applications*, L. Torres, E. Masgrau, and M. A. Lagunas, Eds., Barcelona, Spain, 1990, pp. 673–676, Elsevier.
- [10] S. I. Amari, A. Cichoki, and H. H. Yang, "A new learning algorithm for blind signal separation," in *Neural Information Processing System 8*, Eds. D.S. Tourezyky et. al., 1995, pp. 757–763.
- [11] O. Macchi and E. Moreau, "Self-adaptive source separation using correlated signals and cross-cumulants," in *Proc. Workshop Athos working group*, Girona, Spain, June 1995.
- [12] A. Mansour and C. Jutten, "A direct solution for blind separation of sources," *IEEE Trans. on Signal Processing*, vol. 44, no. 3, pp. 746–748, March 1996.
- [13] M. Gaeta and J. L. Lacoume, "Sources separation without a priori knowledge: the maximum likelihood solution," in *Signal Processing V, Theories and Applications*, L. Torres, E. Masgrau, and M. A. Lagunas, Eds., Barcelona, Spain, 1994, pp. 621–624, Elsevier.
- [14] N. Delfosse and P. Loubaton, "Adaptive blind separation of independent sources: A deflation approach," *Signal Processing*, vol. 45, no. 1, pp. 59–83, July 1995.
- [15] A. Mansour and C. Jutten, "Fourth order criteria for blind separation of sources," *IEEE Trans. on Signal Processing*, vol. 43, no. 8, pp. 2022–2025, August 1995.
- [16] A. Gorokhov and P. Loubaton, "Subspace based techniques for second order blind separation of convolutive mixtures with temporally correlated sources," *IEEE Trans. on Circuits and Systems*, vol. 44, pp. 813–820, September 1997.
- [17] A. Mansour, C. Jutten, and P. Loubaton, "An adaptive subspace algorithm for blind separation of independent sources in convolutive mixture," *IEEE Trans. on Signal Processing*, vol. 48, no. 2, February 2000.
- [18] A. Mansour, C. Jutten, and P. Loubaton, "Subspace method for blind separation of sources and for a convolutive mixture model," in *Signal Processing VIII, Theories and Applications*, Triest, Italy, September 1996, pp. 2081–2084, Elsevier.
- [19] A. Mansour, A. Kardec Barros, and N. Ohnishi, "Subspace adaptive algorithm for blind separation of convolutive mixtures by conjugate gradient method," in *The First International Conference and Exhibition Digital Signal Processing (DSP'98)*, Moscow, Russia, June 30–July 3 1998, pp. I-252–I-260.
- [20] D. Gesbert, P. Duhamel, and S. Mayrargue, "Subspace-based adaptive algorithms for the blind equalization of multichannel fir filters," in *Signal Processing VII, Theories and Applications*, M.J.J. Holt, C.F.N. Cowan, P.M. Grant, and W.A. Sandham, Eds., Edinburgh, Scotland, September 1994, pp. 712–715, Elsevier.
- [21] A. Mansour and N. Ohnishi, "A blind separation algorithm based on subspace approach," in *IEEE-EURASIP Workshop on Nonlinear Signal and Image Processing (NSIP'99)*, Antalya, Turkey, June 20–23 1999, pp. 268–272.
- [22] T. Kailath, *Linear systems*, Prentice Hall, 1980.
- [23] R. Bitmead, S. Kung, B. D. O. Anderson, and T. Kailath, "Greatest common division via generalized Sylvester and Bezout matrices," *IEEE Trans. on Automatic Control*, vol. 23, no. 6, pp. 1043–1047, December 1978.
- [24] A. Mansour, C. Jutten, and P. Loubaton, "Robustesse des hypothèses dans une méthode sous-espace pour la séparation de sources," in *Actes du XVIème colloque GRETSI*, Grenoble, France, septembre 1997, pp. 111–114.
- [25] A. Mansour and N. Ohnishi, "Multichannel blind separation of sources algorithm based on cross-cumulant and the levenberg-marquardt method," *IEEE Trans. on Signal Processing*, vol. 47, no. 11, pp. 3172–3175, November 1999.
- [26] G. Puntinet, C. A. Mansour, and C. Jutten, "Geometrical algorithm for blind separation of sources," in *Actes du XVème colloque GRETSI*, Juan-Les-Pins, France, 18-21 septembre 1995, pp. 273–276.
- [27] A. Prieto, C. G. Puntinet, and B. Prieto, "A neural algorithm for blind separation of sources based on geometric properties," *Signal Processing*, vol. 64, no. 3, pp. 315–331, 1998.
- [28] A. Mansour, A. Kardec Barros, and N. Ohnishi, "Comparison among three estimators for high order statistics," in *Fifth International Conference on Neural Information Processing (ICONIP'98)*, Kitakyushu, Japan, 21-23 October 1998, pp. 899–902.



(a) Performance matrix Perf (b) The criterion is flat around its minima.

Figure 2: Performances and Properties.

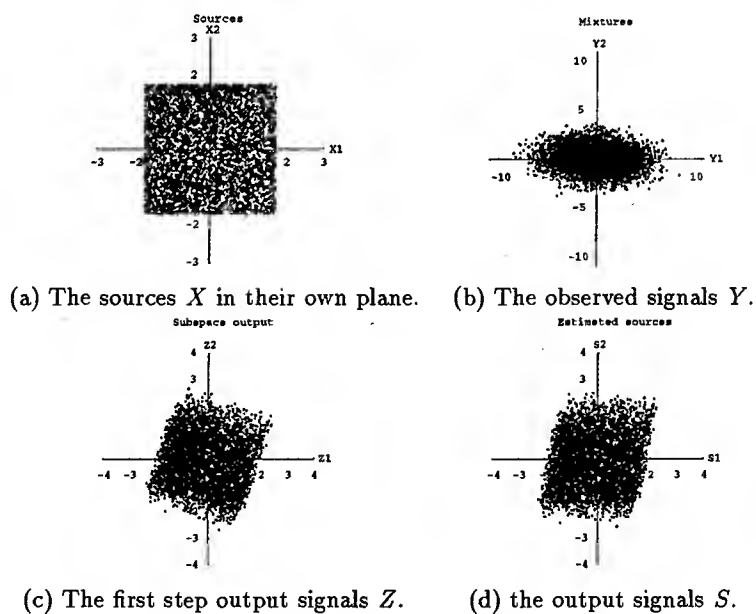


(a) Performance matrix Perf, by only using LMS (b) The Performance matrix, using initialized LMS.



(c) Criterion convergence of the LMS version. (d) Criterion convergence of the initialized LMS version.

Figure 3: Performances and convergence.



(c) The first step output signals Z. (d) the output signals S.

Figure 4: Different signals.

# COMPARATIVE STUDY OF TWO-DIMENSIONAL MAXIMUM LIKELIHOOD AND INTERPOLATED ROOT-MUSIC WITH APPLICATION TO TELESEISMIC SOURCE LOCALIZATION

Pei-Jung Chung\*    Alex B. Gershman\*\*    Johann F. Böhme\*

\*Department of Electrical Engineering and Information Science,  
Ruhr University, D-44780 Bochum, Germany  
pjc,boehme@sth.ruhr-uni-bochum.de

\*\*Department of Electrical and Computer Engineering,  
McMaster University, Hamilton, L8S 4K1 Ontario, Canada  
gershman@ieee.org

## ABSTRACT

We apply the 2-D broadband Maximum Likelihood (ML) and interpolated root-MUSIC methods to estimate the azimuth and velocity parameters of teleseismic events recorded by the GERESS array. A sequential test based on Likelihood Ratios (LR's) is developed for signal detection. Our experimental results show that both methods can provide reliable estimates of signal parameters. However, ML is shown to have better estimation accuracy and robustness than interpolated root-MUSIC at the expense of a higher computational cost.

## 1. INTRODUCTION

The ML and MUSIC techniques are two popular methods in array processing. Numerous theoretical and numerical studies have shown that ML outperforms MUSIC in scenarios with low Signal to Noise Ratios (SNR's), small number of samples, coherent signals, as well as closely spaced sources [1]. However, an enormously high computational cost needed for ML makes this statistically optimal approach in many cases less attractive than MUSIC. Therefore, a crucial issue is how to choose a proper algorithm for a particular application to achieve sufficiently high performance and acceptable computational complexity.

In the present work, we apply broadband ML [2] and 2-D interpolated root-MUSIC [3] to localization of several teleseismic events using the GERESS array real data. A sequential test procedure based on LR's is used to detect signals within the observation interval. Due

---

This work was supported by the German Science Foundation and by the Natural Sciences and Engineering Research Council (NSERC) of Canada.

to complicated propagation effects, there may be more than one *signal phase* arriving at the same time from the same direction. However, different signal phases should differ in their velocities. It is worth noting that the ML method can be directly applied to the broadband Direction Of Arrival (DOA) estimation problem. On the other hand, root-MUSIC should be adapted to the broadband setting, for example, by means of the so-called array interpolation technique [4] allowing to combine the information from different frequencies in a coherent way. In [3] and [6], a high-SNR regional man-made seismic event was analyzed by means of the ML and interpolated root-MUSIC techniques. Both methods provided excellent results in this case. Below, we address a more difficult *teleseismic event* case, which is characterized by much lower SNR's and more complicated propagation phenomena relative to the regional event case. In the teleseismic case, signal detection becomes a very important issue, since it is almost impossible to identify weak signals in seismograms (for example, see Fig. 1 displaying a typical seismogram of teleseismic event).

The experimental results reported in the present paper demonstrate that in the teleseismic case, both ML and interpolated root-MUSIC may be successfully applied to source localization. ML is shown to have better performance and robustness than interpolated root-MUSIC. However, the latter approach enjoys simpler implementation.

## 2. DATA MODEL

Let an array of  $N$  sensors receive  $M$  broadband signals from far-field sources. The 2-D array can be assumed



since the length of the vertical aperture of GERESS is much smaller than that of the horizontal one and is negligible compared to the seismic signal wavelength. The array output  $\underline{x}(t)$  sampled at discrete times  $t = 0, \dots, T-1$  is short-time Fourier-transformed using the so-called Thomson's multitaper technique [7]:

$$\underline{X}_l(\omega) = \frac{1}{\sqrt{T}} \sum_{t=0}^{T-1} w_l(t) \underline{x}(t) e^{-j\omega t}, \quad (l = 0, \dots, L-1), \quad (1)$$

where  $\{w_l(t)\}_{t=0, \dots, T-1}$  is the  $l$ th orthonormal window function.

For sufficiently large  $T$ , the Fourier-transformed data can be approximately expressed as

$$\underline{X}_l(\omega) = \mathbf{H}(\omega) \underline{S}_l(\omega) + \underline{U}_l(\omega), \quad (2)$$

$$\mathbf{H}(\omega) = [\underline{d}_1(\omega), \dots, \underline{d}_M(\omega)], \quad (3)$$

where  $\underline{X}_l(\omega) \in \mathbb{C}^{N \times 1}$ ,  $\mathbf{H}(\omega) \in \mathbb{C}^{N \times M}$ ,  $\underline{S}_l(\omega) \in \mathbb{C}^{M \times 1}$ , and  $\underline{U}_l(\omega) \in \mathbb{C}^{N \times 1}$  are the observation vector, the steering matrix, the vector of signal waveforms, and the vector of sensor noise, respectively. The steering vector  $\underline{d}_m(\omega)$  associated with the  $m$ th signal is given by

$$\underline{d}_m(\omega) = [e^{-i\omega \xi_m^T \mathbf{r}_1}, \dots, e^{-i\omega \xi_m^T \mathbf{r}_N}]^T, \quad (4)$$

where  $\mathbf{r}_n = (x_n, y_n)$  is the coordinate of the  $n$ th sensor. The slowness vector  $\xi_m$  is related to the source azimuth  $\alpha_m$  and the respective velocity  $V_m$  as follows:

$$\xi_m = \frac{1}{V_m} [\cos \alpha_m, \sin \alpha_m]^T. \quad (5)$$

The signal waveforms  $\underline{S}_l(\omega_j)$ , ( $l = 0, \dots, L-1$ ,  $j = 1, \dots, J$ ) are assumed to be deterministic and unknown. From the asymptotic theory of the Fourier transform, it is well-known that  $\underline{X}_l(\omega_j)$ , ( $l = 0, \dots, L-1$ ;  $j = 1, \dots, J$ ) are independent complex Gaussian distributed with the mean  $\mathbf{H}(\omega_j) \underline{S}_l(\omega_j)$  and the covariance matrix  $\nu(\omega_j) \mathbf{I}$  where  $\nu(\omega_j)$  is the sensor noise power at the frequency  $\omega_j$  and  $\mathbf{I}$  is the identity matrix [2]. The problem is to detect the signals and estimate their parameters  $\{\alpha_m, V_m\}$ ,  $m = 1, \dots, M$ .

### 3. WIDEBAND MAXIMUM LIKELIHOOD

Based on the independence and asymptotic gaussianity in the frequency domain, the approximate wideband log-likelihood function can be expressed as [2]

$$l(\underline{\vartheta}) = \sum_{j=1}^J \log \text{tr} \left[ \{\mathbf{I} - \mathbf{P}(\omega_j, \underline{\vartheta})\} \hat{\mathbf{R}}_X(\omega_j) \right], \quad (6)$$

where

$$\underline{\vartheta} = [\xi_1^T, \dots, \xi_M^T]^T \quad (7)$$

denotes the unknown slowness vector,  $\mathbf{P}(\omega_j, \underline{\vartheta})$  is the projection matrix onto the column-space of the matrix  $\mathbf{H}(\omega_j)$ ,

$$\hat{\mathbf{R}}_X(\omega_j) = \frac{1}{L} \sum_{l=0}^{L-1} \underline{X}_l(\omega_j) \underline{X}_l^H(\omega_j) \quad (8)$$

is the sample spectral density matrix, and  $(\cdot)^H$  denotes the Hermitian transpose. The ML estimate  $\hat{\underline{\vartheta}}_{\text{ML}}$  is obtained by minimizing (6) over  $\underline{\vartheta}$ .

### 4. WIDEBAND INTERPOLATED ROOT-MUSIC

In this section, we describe the 2-D extension [3] of the wideband interpolated root-MUSIC algorithm [5] that will be applied for joint estimation of the azimuth and velocity parameters of seismic sources.

Let the 2-D array be divided into two subarrays of  $N_s$  sensors each, denoted as subarrays (a) and (b), respectively. Since the outline of the algorithm is similar for each subarray, in the sequel we consider only the subarray (a). Its observation vector can be modeled as

$$\underline{X}_{l,a}(\omega) = \mathbf{H}_a(\omega) \underline{S}_l(\omega) + \underline{U}_{l,a}(\omega). \quad (9)$$

This subarray will be used for interpolation of the set of  $J$  virtual ULA's with the interelement spacings  $d_c \omega_c / \omega_j$  ( $j = 1, \dots, J$ ), where  $\omega_c$  is the central frequency, and  $d_c$  is the interelement spacing of the virtual ULA at  $\omega_c$ . To obtain the same array manifold for each frequency, the interpolation matrices  $\mathbf{B}_j$  can be designed in a regular way [4]. The coherently averaged covariance matrix can be obtained as

$$\tilde{\mathbf{R}}_a = \frac{1}{J} \sum_{j=1}^J \mathbf{B}_j^H \hat{\mathbf{R}}_a \mathbf{B}_j, \quad (10)$$

where

$$\hat{\mathbf{R}}_a = \frac{1}{L} \sum_{l=0}^{L-1} \underline{X}_{l,a}(\omega) \underline{X}_{l,a}^H(\omega). \quad (11)$$

The noise covariance matrix after the coherent processing can be computed as

$$\hat{\mathbf{Q}} = \frac{1}{J} \sum_{j=1}^J \hat{\nu}(\omega_j) \mathbf{B}_j^H \mathbf{B}_j, \quad (12)$$

where  $\hat{\nu}(\omega_j)$  is some estimate of sensor noise at the frequency  $\omega_j$ . The matrix

$$\bar{\mathbf{R}}_a = \hat{\mathbf{Q}}^{-1/2} \tilde{\mathbf{R}}_a \hat{\mathbf{Q}}^{-1/2} \quad (13)$$

is the spectral density matrix after prewhitening. The eigendecomposition of this matrix yields

$$\bar{\mathbf{R}}_a = \mathbf{U}_S \mathbf{\Lambda}_S \mathbf{U}_S^H + \mathbf{U}_N \mathbf{\Lambda}_N \mathbf{U}_N^H, \quad (14)$$

where the matrices  $\mathbf{U}_S$  and  $\mathbf{U}_N$  contain the signal- and noise-subspace eigenvectors, respectively. In turn, the diagonal matrices  $\mathbf{\Lambda}_S$  and  $\mathbf{\Lambda}_N$  contain the signal- and noise-subspace eigenvalues, respectively.

The root-MUSIC polynomial has the form

$$D_a(z) = \underline{d}^T(1/z) \hat{\mathbf{Q}}^{-1/2} \mathbf{U}_N \mathbf{U}_N^H \hat{\mathbf{Q}}^{-1/2} \underline{d}(z), \quad (15)$$

where  $\underline{d}(z) = [1, z^{-1}, \dots, z^{N-1}]^T$ . Let  $\{z_{a,1}, \dots, z_{a,M}\}$  denote the  $M$  signal roots of (15), which are sorted based on their proximity to the unit circle. Similarly, we can find  $M$  signal roots  $\{z_{b,1}, \dots, z_{b,M}\}$  for subarray (b). Combining the results from these two virtual subarrays, we can find  $M^2$  candidate estimates of  $\underline{\xi}$  by solving the system

$$\begin{aligned} \Delta x_a \xi_x + \Delta y_a \xi_y &= \arg \frac{z_{a,i}}{\omega_c}, \\ \Delta x_b \xi_x + \Delta y_b \xi_y &= \arg \frac{z_{b,k}}{\omega_c} \end{aligned} \quad (16)$$

for  $i, k = 1, \dots, M$ , where  $\Delta x_a$ ,  $\Delta y_a$ ,  $\Delta x_b$ , and  $\Delta y_b$  define the interelement spacings of the virtual arrays (a) and (b), respectively. The final estimate of  $\underline{\xi}$  is then obtained by selecting the  $M$  pairs  $(\hat{\xi}_x, \hat{\xi}_y)$  which correspond to the maximal values of 2-D MUSIC spectral function. The estimates of azimuth and velocity  $\hat{\underline{v}}_{\text{MUSIC}}$  can be obtained from these  $M$  pairs using (5).

## 5. LIKELIHOOD RATIO TEST

In this section, we develop a sequential LR-based test for detecting the number of signals. Let  $m$  denote the hypothetical number of signals. In each step, the detection problem can be formulated as testing the hypothesis  $\mathcal{H}_m$  against the alternative  $\mathcal{A}_m$ :

$$\begin{aligned} \mathcal{H}_m & \quad m \text{ signals are present,} \\ \mathcal{A}_m & \quad \text{more than } m \text{ signals are present.} \end{aligned}$$

Starting from  $m = 0$ , this test should be performed stepwise and then stopped once the hypothesis  $\mathcal{H}_m$  becomes accepted. Applying LR principle, we obtain the following test statistic in the  $m$ th step [2]:

$$t_m = \frac{1}{J} \sum_{j=1}^J \log(1 + \frac{n_1}{n_2} F_m(\omega_j)) \geq t_\alpha, \quad (17)$$

where

$$F_m(\omega) = \frac{n_2 \text{tr} \left[ \left\{ \mathbf{P}_{m+1}(\omega, \hat{\underline{v}}_{\text{ML}}^{m+1}) - \mathbf{P}_m(\omega, \hat{\underline{v}}_{\text{ML}}^m) \right\} \hat{\mathbf{R}}_X(\omega) \right]}{n_1 \text{tr} \left[ \left\{ \mathbf{I} - \mathbf{P}_{m+1}(\omega, \hat{\underline{v}}_{\text{ML}}^{m+1}) \right\} \hat{\mathbf{R}}_X(\omega) \right]}, \quad (18)$$

$$n_1 = L(2m + 4), n_2 = L(2N - 2), \quad (19)$$

and  $\hat{\underline{v}}_{\text{ML}}^m \in \mathbb{R}^{2m}$  is the ML estimate of the signal parameter vector. If  $t_m$  exceeds the test threshold  $t_\alpha$ , the hypothesis will be rejected. The quantity calculated by  $F_m(\omega)$  can be interpreted as an estimate of the increase in SNR when adding the  $(m+1)$ th signal. To be detected, the power of  $(m+1)$ th signal must be sufficiently high compared to the noise power. Under the hypothesis  $\mathcal{H}_m$ , the value  $F_m(\omega_k)$  is approximately centrally  $F$ -distributed with the degrees of freedom  $n_1$  and  $n_2$ . The threshold  $t_\alpha$  is determined by the Cornish-Fisher expansion with a good accuracy [8]-[9]. Note that the LR test can be easily implemented if the corresponding ML estimates are available.

## 6. REAL DATA PROCESSING

In this section, we apply the developed techniques to real data processing. These data were recorded by the GERESS array located in the Bavarian Forest, Germany. Details about this array can be found in [10]. Two teleseismic events (earthquakes) which occurred on February 13, 1993 in the Eastern Mediterranean and on February 26, 1996 in the Middle East, respectively, were selected for our analysis. The latter event is contaminated by a smaller pre-shock, located about 37 km from the main event. More information about the selected events is collected in Table 1.

Array output was sampled with  $f_s = 40$  Hz. For each data set, we used a sliding window with the length of 3.2 s and the shift of 0.5 s. The total of seven frequency bins between 0.9 and 3.1 Hz have been used. Two independent virtual ULA sets have been employed for the interpolated root-MUSIC algorithm with the central frequency  $f_c = 2.2$  Hz. The spectral density matrix  $\hat{\mathbf{R}}_X(\omega_j)$  has been estimated using  $L = 3$  Thomson's windows which roughly correspond to 3 independent snapshots. The sequential detection procedure kept the test level  $\alpha = 0.033$  constant in each step. Theoretical slowness values have been derived from AK135 earth model [11].

The results obtained from the weak event analysis are shown in Figs. 1 and 2. Typical seismometer outputs are plotted in the first subplot of these figures. The second subplot shows the output of the LR-based detector which was used in conjunction with both techniques to provide their adequate comparison. Apparently, the P-phases are detected with a good time resolution while the S-phases (traveling with lower velocity) are not detected at all. Some false alarms can be observed. The ML estimates for the back-azimuth and velocity are well concentrated around their theoretical values. The estimates obtained from 2-D interpolated

Table 1: Event List from NEIC.

time h.m.s	lat deg N	long deg E	dist deg	az deg	mag mb	locat
03:42:53	34.43	24.81	16.60	146.1	3.7	Crete
07:17:08	28.87	34.48	25.54	133.8	4.0	Gulf Aqaba
07:17:28	28.73	34.82	25.81	133.4	5.0	Gulf Aqaba

root-MUSIC show higher variances. Interestingly, both methods provide better results for the azimuth than the velocity. Such a relatively poor performance of velocity estimates may be explained by quite a limited aperture length of GERESS.

In Figs. 3 and 4, another event is analyzed. It contains two seismic sources of moderate scales originating from the same location but at slightly different times (see Table 1). In this data set, a stronger event follows shortly after a weak event. In particular, such a situation is of great importance when monitoring nuclear explosions. Due to high SNR's, the signals can be correctly detected during the whole analysis interval. One signal is detected at about 30th second when waves from the first earthquake arrive the array. At 57th second, the LR test shows two signals, corresponding to the case when the superimposing waves from the first and second seismic sources both arrive the array. During the period from 300th to 360th second (the so-called S-phases), similar detection results can be observed as well. The signals detected from the beginning of the analysis up to 16th second could be interpreted as false alarms or another weak event. The estimates of the azimuth and velocity shown in subplots 3 and 4 illustrate that the ML technique has better robustness and lower variance than the 2-D interpolated root-MUSIC technique. Note that the performance of the latter method is not much better in the strong event case than in the weak event one, since the interpolation errors become more critical at high SNR's. Similarly to the previous example, both methods show better azimuth estimation performance relative to that of velocity estimation.

## 7. CONCLUSIONS

We compared the performances of wideband ML and interpolated root-MUSIC algorithms by processing weak and strong teleseismic events recorded by the GERESS array. Our results show that ML has better estimation accuracy and robustness relative to root-MUSIC. Another advantage of ML is that the application of the LR test for detecting the number of signals is straightforward. However, the enormous computational cost

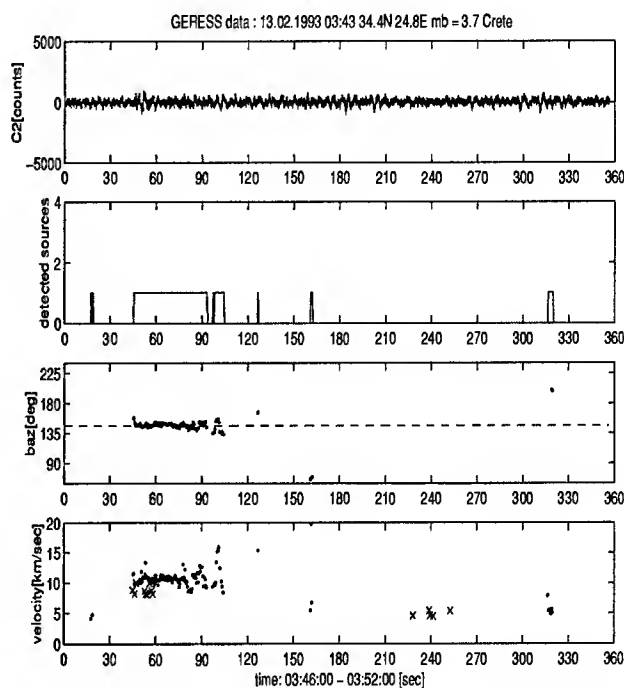


Figure 1: Wideband ML, first event. "—": theoretical values for back-azimuth, "x": theoretical values for velocity.

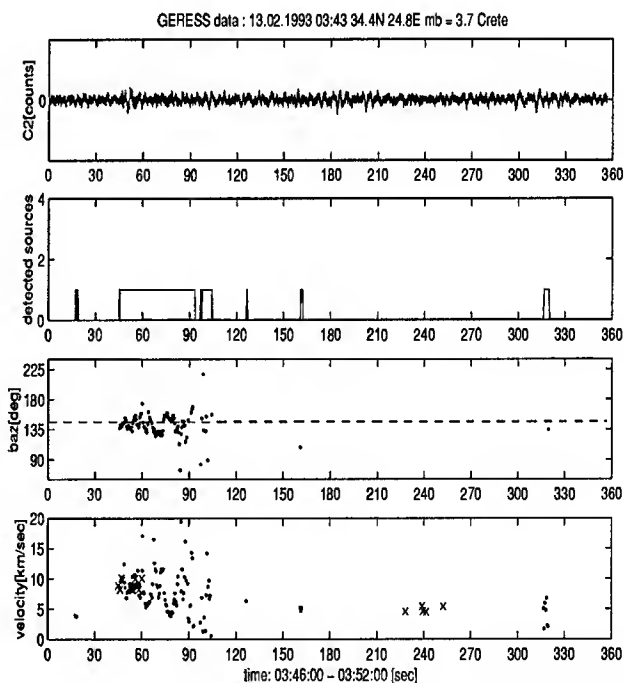


Figure 2: Wideband interpolated root-MUSIC, first event. "—": theoretical values for back-azimuth, "x": theoretical values for velocity.

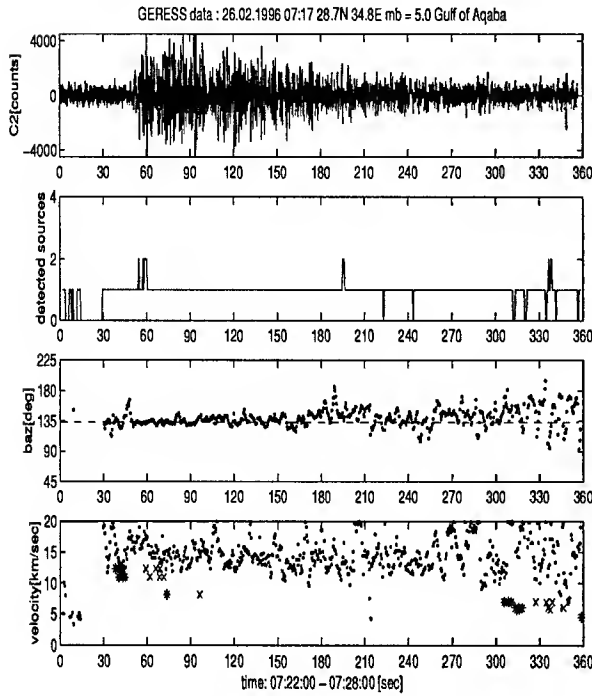


Figure 3: Wideband ML, second event. "—": theoretical values for back-azimuth, "x": theoretical values for velocity of the main event, "\*" : theoretical values for velocity of the pre-shock.

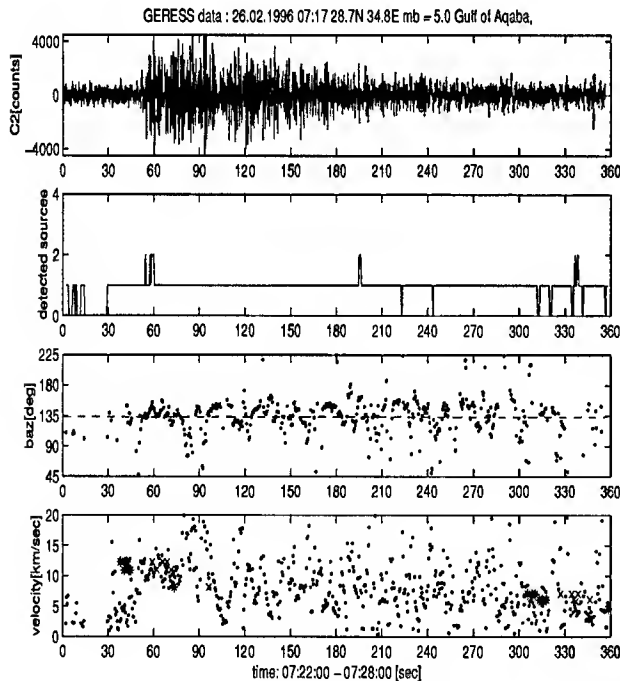


Figure 4: Wideband interpolated root-MUSIC, second event. "—": theoretical values for back-azimuth, "x": theoretical values for velocity of the main event, "\*" : theoretical values for velocity of the pre-shock.

associated with the ML technique may be critical in practical applications.

## REFERENCES

- [1] J.F. Böhme, "Advances in spectrum analysis and array processing," in *Array Processing*, Haykin, S., Editor, Prentice Hall, pp. 1-63, 1991.
- [2] J.F. Böhme, "Statistical array signal processing of measured sonar and seismic data," in *Proc. SPIE 2563: Advanced Signal Processing Algorithms*, San Diego, CA, July 1995, pp. 2-20.
- [3] D.V. Sidorovich and A.B. Gershman, "Two-dimensional wideband interpolated root-MUSIC applied to measured seismic data," *IEEE Trans. Signal Processing*, vol. 46, pp. 2263-2267, Aug. 1998.
- [4] B. Friedlander, "The root-MUSIC algorithm for direction finding with interpolated arrays," *Signal Processing*, vol. 30, pp. 15-29, Jan. 1993.
- [5] B. Friedlander and A.J. Weiss, "Direction finding for wideband signals using an interpolated array," *IEEE Trans. Signal Processing*, vol. 41, pp. 1618-1634, Apr. 1993.
- [6] D.V. Sidorovich, C.F. Mecklenbräuker, and J.F. Böhme, "Sequential test and parameter estimation for array processing of seismic data," in *Proc. 8th IEEE Workshop Stat. Signal Array Processing*, Corfu, Greece, June 1996, pp. 256-259.
- [7] D.J. Thomson, "Spectrum estimation and harmonic analysis," *Proc. IEEE*, vol. 70, pp. 1055-1096, Sep. 1982.
- [8] P. Hall, *The Bootstrap and Edgeworth Expansion*, Springer-Verlag, NY, 1992.
- [9] C.F. Mecklenbräuker, P. Gerstoft, J.F. Böhme, and P.-J. Chung, "Hypothesis testing for geoaoustic environmental models using likelihood ratio," *JASA*, vol. 105, pp. 1738-1748, March 1999.
- [10] H.P. Harjes, "Design and siting of a new regional array in Central Europe," *Bull. Seism. Soc. Am.*, vol. 80B, pp. 1801-1817, June 1990.
- [11] B. Kennett, E.R. Engdahl, and R. Buland, "Constraints on seismic velocities in the Earth from traveltimes," *Geophys. J. Int.*, vol. 122, pp. 108-124, 1995.

# BOUNDS ON UNCALIBRATED ARRAY SIGNAL PROCESSING

Brian M. Sadler

Army Research Laboratory  
Adelphi, MD 20783  
bsadler@arl.mil

Richard J. Kozick

Bucknell University  
Lewisburg, PA 17837  
kozick@bucknell.edu

## ABSTRACT

Deterministic constrained Cramér-Rao bounds (CRBs) are developed for general linear forms in additive white Gaussian noise. The linear form describes a variety of array processing cases, including narrow band sources with a calibrated array, the uncalibrated array cases of instantaneous linear mixing and convolutive mixing, and space-time coding scenarios with multiple transmit and receive antennas. We employ the constrained CRB formulation of Stoica and Ng, allowing the incorporation of side information into the bounds. This provides a framework for a large variety of scenarios, including semi-blind, constant modulus, known moments or cumulants, and others. The CRBs establish bounds on blind estimation of sources using an uncalibrated array, and facilitates comparison of calibrated and uncalibrated arrays when side information is exploited.

## 1. INTRODUCTION: MODEL

Consider the additive noise linear model

$$\mathbf{x}_t = H\mathbf{s}_t + \mathbf{v}_t, \quad t = 1, \dots, N, \quad (1)$$

where  $\mathbf{x}_t$  is  $l \times 1$  and  $H$  is  $l \times k$ . The elements of the  $k \times 1$  signal vector will be denoted by  $\mathbf{s}_t = [s_1(t), \dots, s_k(t)]^T$ . We use the notation superscript  $T, *, H$  for transpose, conjugate, and conjugate transpose, respectively, with complex numbers denoted  $c = \bar{c} + j\tilde{c}$ . The noise  $\mathbf{v}_t$  is assumed complex white Gaussian, with variance  $\sigma^2$ . The model (1) underlies many array processing and single-sensor scenarios.

In the narrow band calibrated array case ( $l$  sensors and  $k$  sources),  $H = A(\theta) \cdot \alpha$  is of known parametric form with respect to the source bearings. Here  $A(\theta)$  is the array manifold matrix, and  $\alpha = \text{diag}(\alpha_1, \dots, \alpha_k)$  contains complex constants  $\alpha_i$  that model the channel attenuation for the  $i$ th source. Constrained bounds are developed for this case in [1, 2].

In this paper we are interested in the general case when  $H$  is unknown. This arises in the uncalibrated array cases of instantaneous linear mixing and convolutive mixing, and the space-time transmit diversity case with arrays for both transmission and reception. An uncalibrated array may have unknown sensor placement, phase mis-matching, and so on. In such cases blind methods may be used to separate and estimate source waveforms without estimating the source bearings. Performance bounds are not straightforward due to the lack of regularity in the Fisher information matrix (FIM) associated with (1) in the uncalibrated case.

We develop CRBs for these cases using the constrained CRB methodology of Gorman/Hero and Stoica/Ng [3]. The constraints arise due to side information such as constant modulus sources, constraints on the structure and elements of  $H$ , and semi-blind sources (some known signal values). Examples are given comparing calibrated and uncalibrated array CRBs. A space-time coding example is also presented.

## 2. FIM & CONSTRAINED CRBS

Forming the  $lN \times 1$  supervector  $\mathbf{x} = [\mathbf{x}_1^T, \dots, \mathbf{x}_N^T]^T$ , then  $\mathbf{x} \sim \text{CN}(\mu_x, R_x = \sigma^2 \mathbf{I}_{lN \times lN})$ , where

$$\mu_x = E[\mathbf{x}] = [\mu_1^T, \dots, \mu_N^T]^T, \quad \mu_t = H\mathbf{s}_t. \quad (2)$$

Thus we have a multivariate complex normal process with deterministic time-varying mean  $H\mathbf{s}_t$ . We define the data matrix and the columns of  $H$  as

$$S = [\mathbf{s}_1, \dots, \mathbf{s}_N]_{k \times N}, \quad H = [\mathbf{h}_1, \dots, \mathbf{h}_k]. \quad (3)$$

We write the unknown deterministic parameters in a real vector of length  $2lk + 2kN$ , given by

$$\Theta = [\Theta_H^T, \Theta_S^T]^T, \quad \Theta_H = [\tilde{\mathbf{h}}_1^T, \tilde{\mathbf{h}}_1^T, \dots, \tilde{\mathbf{h}}_k^T, \tilde{\mathbf{h}}_k^T]^T, \\ \Theta_S = [\tilde{\mathbf{s}}_1^T, \tilde{\mathbf{s}}_1^T, \dots, \tilde{\mathbf{s}}_N^T, \tilde{\mathbf{s}}_N^T]^T. \quad (4)$$

Note that  $\sigma^2$  decouples from the other parameters, and so it is omitted.

The FIM  $J$  for  $\Theta$  is obtained from

$$[J(\Theta)]_{ij} = \frac{2}{\sigma^2} \text{Re} \left[ \sum_{t=1}^N \frac{\partial \mu_t^H}{\partial \Theta_i} \frac{\partial \mu_t}{\partial \Theta_j} \right]. \quad (5)$$

Partitioning  $J$  we write,

$$J = \begin{bmatrix} J_H & J_{HS} \\ J_{SH} & J_S \end{bmatrix}, \quad (6)$$

with elements described next. Define the  $2k \times 2k$  matrix

$$J_0 = \begin{bmatrix} H^H H & jH^H H \\ -jH^H H & H^H H \end{bmatrix}, \quad (7)$$

then  $J_S$  is given by the block-diagonal  $2kN \times 2kN$  matrix

$$J_S = \frac{2}{\sigma^2} \text{Re}\{\text{diag}(J_0, \dots, J_0)\}, \quad (8)$$

where  $J_0$  repeats  $N$  times.

$J_H$  may be written as follows,

$$P = [SS^H]^*_{k \times k} \quad (9)$$

$$B_{mn} = \begin{bmatrix} [P]_{mn} & j[P]_{mn} \\ -j[P]_{mn} & [P]_{mn} \end{bmatrix} \otimes I_{l \times l} \quad (10)$$

$$J_H = \frac{2}{\sigma^2} \text{Re} \begin{bmatrix} B_{11} & \cdots & B_{1k} \\ \vdots & \ddots & \vdots \\ B_{k1} & \cdots & B_{kk} \end{bmatrix}_{2lk \times 2lk}, \quad (11)$$

where  $\otimes$  denotes Kronecker product.

Next we consider the cross-terms in the FIM,  $J_{HS}$  and  $J_{SH}$ . It can be shown that

$$L_{mn} = \begin{bmatrix} [S]_{mn}^* & j[S]_{mn}^* \\ -j[S]_{mn}^* & [S]_{mn}^* \end{bmatrix} \otimes H \quad (12)$$

$$J_{HS} = \frac{2}{\sigma^2} \text{Re} \begin{bmatrix} L_{11} & \cdots & L_{1N} \\ \vdots & \ddots & \vdots \\ L_{k1} & \cdots & L_{kN} \end{bmatrix}_{2lk \times 2kN} = J_{SH}^T. \quad (13)$$

As noted, the FIM  $J$  is generally not invertible because the model parameters are not identifiable, and so no unbiased estimator for  $\Theta$  exists. However, it is possible to achieve identifiability, and then regularity of the FIM, by establishing constraints on  $\Theta$ . We establish  $K$  equality constraints on elements of  $\Theta$ , where  $K < \dim(\Theta)$ . The constraints have the form  $f_i(\Theta) = 0$  for  $i = 1, \dots, K$ . Define a  $K \times 1$  constraint vector  $f(\Theta)$ , and a corresponding  $K \times M$  gradient matrix

$$F(\Theta) = \frac{\partial f(\Theta)}{\partial \Theta} \quad (14)$$

with elements  $[F(\Theta)]_{i,m} = \partial f_i(\Theta) / \partial [\Theta]_m$ . The gradient matrix  $F(\Theta)$  is assumed to have full row rank  $K$  for any  $\Theta$  satisfying the constraints  $f_1(\Theta), \dots, f_K(\Theta)$ . Then, the constrained CRB is obtained via (Thrm. 1 of [3])

$$E[(\hat{\Theta} - \Theta)^T(\hat{\Theta} - \Theta)] \geq U(U^T J U)^{-1} U^T. \quad (15)$$

$J$  is the unconstrained FIM from (5), and  $U$  is an orthonormal basis for the null space of  $F(\Theta)$ , i.e.,  $FU = 0$  and  $U^T U = I$ . Note that  $U$  is a function of the constraints only.

Examples of source constraints of interest include constant modulus (CM) sources, known source cumulant or kurtosis, and semi-blind sources (some known source samples). Constraints may also be placed on  $H$ , such as limiting the norm of  $H$ . Together, sufficient constraints may be found to insure information regularity. These provide CRBs on symbol estimation in blind source separation scenarios that exploit source features such as CM. We may also compare bounds on source estimation for both calibrated and uncalibrated arrays using the results of [1, 2], where we have established CRBs on bearing, symbol, and channel estimation for calibrated arrays with side information.

### 3. EXAMPLES IN ARRAY PROCESSING

We use the constrained CRB formulation to gain insight into the following questions.

1. Which provides more accurate signal copy: an uncalibrated array (unknown  $H$  matrix in (1)) with CM signals, or a calibrated array ( $H = A(\theta) \cdot \alpha$ ) with unconstrained signals?
2. Algorithms for blind beamforming with uncalibrated arrays often exploit independence between the signals and non-Gaussianity as characterized by the kurtosis [4, 5, 6]. What is the relative value of these constraints when compared with the CM constraint for CM signals? Do the CRBs based on kurtosis constraints imply any difference in separability of CM and QAM signals?

We generate observations  $x_1, \dots, x_N$  in (1) using a complex narrowband array model in which  $H = A(\theta) \cdot \alpha$ , where  $A(\theta) = [a(\theta_1), \dots, a(\theta_k)]$  is the array response matrix,  $\theta = [\theta_1, \dots, \theta_k]^T$  are the source angles of arrival (AOAs),  $a(\theta_i)$  is the array manifold, and  $\alpha = \text{diag}\{\alpha_1, \dots, \alpha_k\}$  is a diagonal complex channel gain matrix. We consider a uniform linear array (ULA) with omnidirectional sensors and half-wavelength spacing, so the array manifold elements are  $[a(\theta)]_m = \exp[j\pi(m-1)\sin\theta]$ ,  $m = 1, \dots, l$ .

Consider a particular ULA with  $l = 5$  sensors and  $k = 2$  sources with AOAs  $\theta_1 = 0^\circ$  and  $\theta_2$  varying from  $1^\circ$  to  $30^\circ$ , where the AOAs are measured with respect to the array broadside. The noise variance is  $\sigma^2 = 1$ , and the number of time samples is  $N = 100$ . The complex amplitudes  $\alpha_1$  and  $\alpha_2$  are generated with phase shifts  $\angle \alpha_1 = \frac{\pi}{6}$  and  $\angle \alpha_2 = -\frac{\pi}{4}$  rad. The amplitudes  $|\alpha_1|$  and  $|\alpha_2|$  are chosen to achieve a desired sample SNR, defined as  $\text{SNR}_i = |\alpha_i|^2 \hat{C}_{21}(i) / \sigma^2$  where the sample variance of signal  $i$  is  $\hat{C}_{21}(i) = (1/N) \sum_{t=1}^N |s_i(t)|^2$ .  $\text{SNR}_1$  is fixed at 10 dB, while  $\text{SNR}_2$  is evaluated at 5, 10, and 15 dB. One beamwidth for the array is  $23.6^\circ$  at broadside.

#### 3.1. Calibrated vs. uncalibrated arrays

The constrained CRB for a calibrated array in which  $H$  has the structure  $A(\theta) \cdot \alpha$  is presented elsewhere [2]. Here we compare the calibrated array CRBs with the uncalibrated array CRBs outlined in the previous section (5), (15). The signal vectors  $s_1, \dots, s_N$  are 8-PSK waveforms with unit modulus  $|s_i(t)| = 1$  and phase rotation such that  $s_1 = [1, \dots, 1]^T$ . For the case of unconstrained mixing matrix  $H$ , it is known [7] that the CM signal constraint and the specified phase rotation are sufficient to uniquely identify  $H$  and the signal phases  $\angle s_i(t)$ . For the case of a calibrated ULA, it is well-known that the AOAs  $\theta$  and signals  $s_i(t)$  are identifiable with no signal constraints ("blind" signals).

Figure 1(a) contains the mean CRB on the signal phase parameters  $\angle s_i(2), \dots, \angle s_i(N)$  for sources  $i = 1, 2$  and various constraints on the structure of  $H$  and the signals  $s_i$ . Note that as the source spacing decreases to less than one beamwidth, the constraints of CM signals with an uncalibrated array (unknown  $H$ ) potentially provide more accuracy in signal phase than a calibrated array with blind signals. Further, the  $\circ$  and  $\times$  symbols are coincident on the plots. So for CM signals, a calibrated array provides negligible improvement in signal phase accuracy compared with an uncalibrated array that places no constraints on  $H$ . This example adds further testament to the well-known power of the CM signal constraint for signal separation.

### 3.2. Uncalibrated array and moment constraints

The following constraints on the signal moments are common in blind beamforming algorithms, e.g., [4]–[6]:

$$\frac{1}{N} S S^H = \text{known matrix, typically } I \quad (16)$$

$$\hat{C}_{20}(i) = \frac{1}{N} \sum_{t=1}^N s_i(t)^2 \text{ is known, } i = 1, \dots, k \quad (17)$$

$$\hat{m}_{42}(i) = \frac{1}{N} \sum_{t=1}^N |s_i(t)|^4 \text{ is known, } i = 1, \dots, k. \quad (18)$$

These are sample moments and not expectations. Note (16) expresses that the signals are uncorrelated, and the diagonal elements of (16) constrain the signal sample variances  $\hat{C}_{21}(i) = 1$ . Then (16)–(18) imply that the signal sample kurtoses  $\hat{C}_{42}(i) = \hat{m}_{42} - |\hat{C}_{20}(i)|^2 - 2\hat{C}_{21}(i)^2$  are known. We will refer to (16)–(18) as “moment constraints,” and we further assume that the first sample of each source signal  $s_1$  is known in order to obtain an invertible constrained FIM. We consider two types of signals: both source signals are 8-PSK (CM), and both source signals are 64 QAM.

Figures 1(b)–(d) contain constrained CRBs for this scenario. For the CM signals, we have also included on the plots the CRBs based on the CM signal constraints  $|s_i(t)| = 1$ ,  $t = 2, \dots, N$ ,  $i = 1, \dots, k$ . The CM signal constraints are exploited by some blind beamforming algorithms, e.g., ACMA [7].

Figure 1(b) contains mean CRBs for the elements of the  $H$  matrix. In the bottom panel in which source 2 is strong ( $\text{SNR}_2 = 15$  dB), the moment constraints and the CM constraints yield about the same CRBs for most values of  $\theta_2$ . In difficult scenarios where the sources become very closely spaced (less than  $10^\circ$ ), the CM signal constraint becomes more informative than the moment constraints. Similar behavior is exhibited in the top panel of Figure 1(b): source 2 is weaker ( $\text{SNR}_2 = 5$  dB), so the CM constraints are more informative than the moment constraints over a larger range of AOA spacings. Note also that if only moment constraints are used, QAM signals provide lower CRBs on  $H$  than CM signals for this case.

Mean CRBs for estimation of the signals  $s_2, \dots, s_N$  are shown in Figures 1(c) and (d). Source 2 is weaker in Figure 1(c) than in Figure 1(d), and we have also included the CRBs for signal estimation when the  $H$  matrix is known perfectly (marked with boxes) but no signal constraints are applied (the blind, calibrated case). In difficult situations of low SNR and closely-spaced sources, exploiting the CM property provides the potential for better performance compared with the moment constraints. Note that the CRBs for signal moment constraints and unconstrained  $H$  are approximately equal to the CRBs for known mixing matrix  $H$  and unconstrained signals, which is similar to our observations about calibrated vs. uncalibrated arrays in Section 3.1.

## 4. SPACE-TIME CODING

Space-time coding employs multiple antennas on transmit and receive [8]. In the flat fading case the model of (1) arises with  $k$  transmit and  $l$  receive antennas, where  $s_t$  is the  $k \times 1$

code vector transmitted by the  $k$  antennas at time  $t$ , and  $[H]_{ij}$  is the complex fading channel gain from the  $j$ th transmit antenna to the  $i$ th receive antenna. The independent Rayleigh fading model corresponds to the  $[H]_{ij}$  being independent, complex, Gaussian random variables with zero mean and unit variance. Suppose that the signal constellation is assumed to have average energy equal to one, and let  $E_s$  denote the total energy transmitted from all  $k$  antennas per symbol. Then we use  $\sqrt{E_s/k} \cdot H$  in the model (1), yielding an average SNR per receive antenna equal to  $E_s/\sigma^2$  for independent, flat, Rayleigh fading channels.

The model (1) assumes that the fading coefficients  $[H]_{ij}$  are constant over the block of  $N$  symbol times. The constrained CRBs developed in this paper assume that  $H \cdot s_t$  in (1) is deterministic, so constrained CRBs may be computed for a particular realization of the fading matrix  $H$ . In the example presented next, we average the CRBs from multiple independent realizations of  $H$  to investigate the diversity gain that results from various constraints.

### 4.1. Constraints

As an example, consider the two-transmit antenna space-time coding scheme in [9]. The code in [9] for  $k = 2$  transmitters can be expressed via the signal constraints

$$s_{t+1} = P s_t^*, \quad t = 1, 3, \dots, N-1 \quad (N \text{ even}) \quad (19)$$

$$\text{where } P = \begin{bmatrix} 0 & -1 \\ 1 & 0 \end{bmatrix}, \quad (20)$$

so a total of two complex symbols are encoded in  $s_t$  and  $s_{t+1}$ . Sampling at the symbol rate is assumed, and this encoding leads to a simple linear receiver structure for maximum likelihood (ML) symbol detection. The ML detector requires knowledge of the channel matrix  $H$ , and training samples are suggested in [9] for estimation of  $H$ . We investigate bounds on estimation of the signals  $s_t$  in the space-time coding context with  $T < N$  training symbols (semi-blind), the code (19), and other constraints including CM signals and known  $H$  matrix.

Suppose that the first  $T$  symbols  $s_1, \dots, s_T$  transmitted from both antennas are known, and assume that  $T < N$  with  $T$  and  $N$  even. Then the gradient matrix (14) corresponding to the  $T$  training symbols (semi-blind) and the space-time code (19) for samples  $T+1, \dots, N$  has the form

$$F_1 = \begin{bmatrix} 0_{k(T+N) \times 2lk} & \begin{matrix} I_{2kT \times 2kT} & F_0 & & \\ & F_0 & \ddots & \\ & & & F_0 \end{matrix} \end{bmatrix} \quad (21)$$

where  $F_0$  repeats  $(N-T)/2$  times and equals

$$F_0 = \begin{bmatrix} I_{4 \times 4} & P & 0_{2 \times 2} \\ 0_{2 \times 2} & -P & \end{bmatrix}. \quad (22)$$

The constraints characterized by (21) will be denoted ‘SEMI-BLIND & S-T CODE’ in the example below. We also consider other combinations of constraints. ‘SEMI-BLIND’ includes training symbols  $s_1, \dots, s_T$  that could be used to jointly estimate  $H$  and the unknown signals  $s_{T+1}, \dots, s_N$ , but the space-time code is not exploited. We can apply the

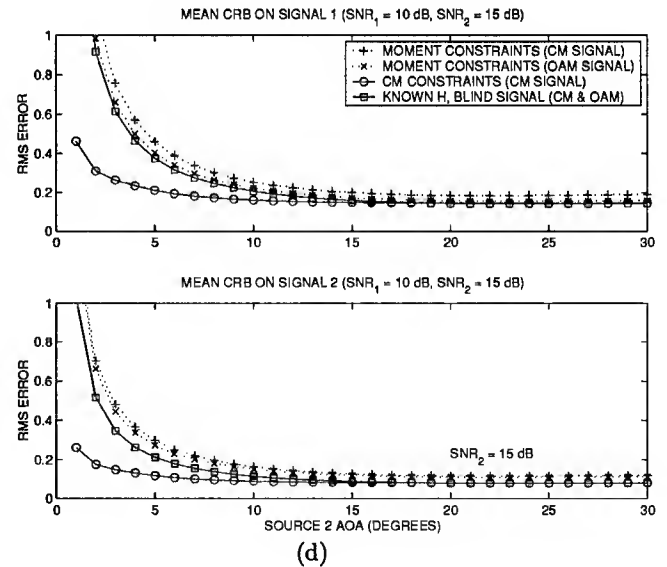
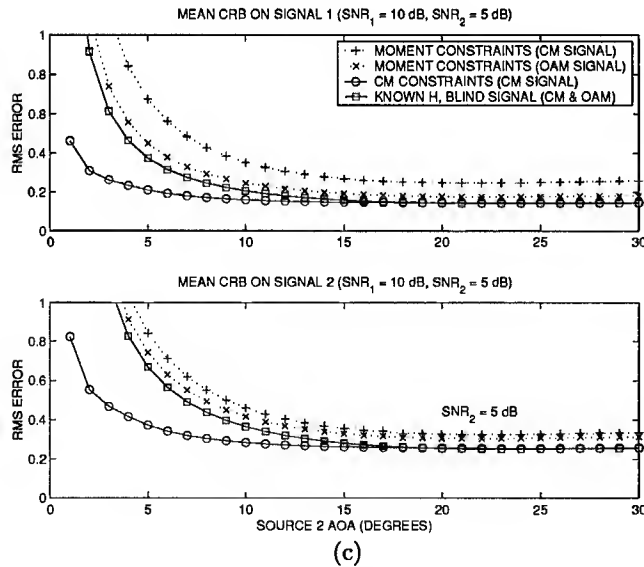
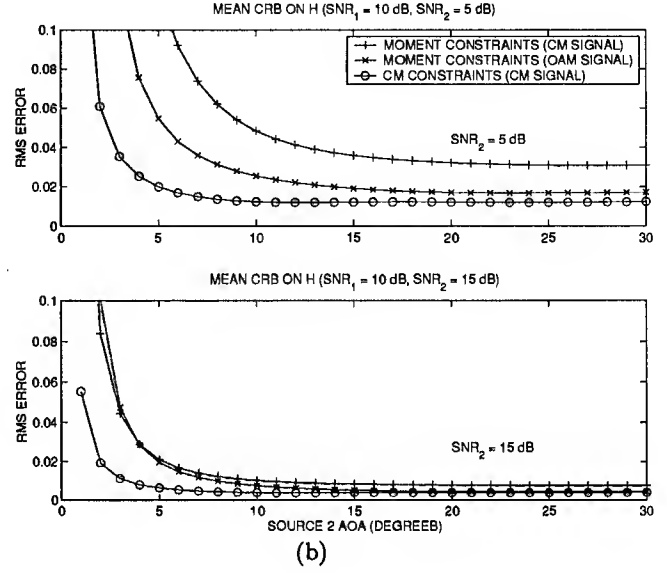
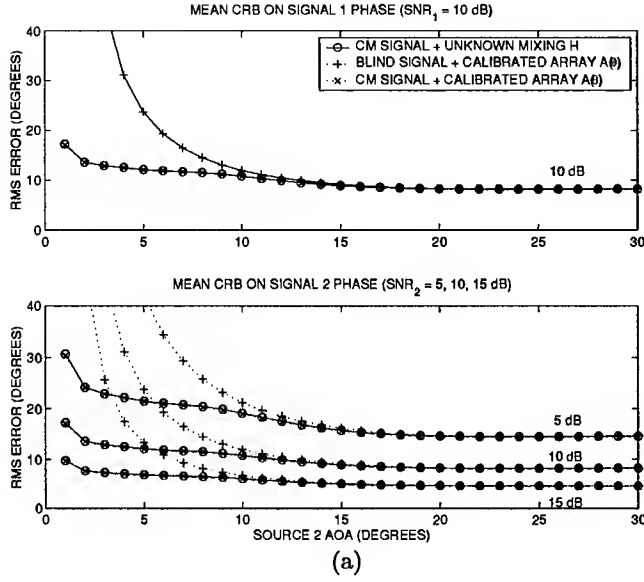


Figure 1: Source 1 bearing is fixed at  $\theta_1 = 0^\circ$ , source 2 bearing is varied on  $[1^\circ, 30^\circ]$ . (a) Uncalibrated vs. calibrated arrays: CRB on signal phase estimation for 8-PSK signals. (b) Mean CRB on elements of  $H$  matrix for 8-PSK (CM) signals and 64-QAM signals for various constraints. (c)-(d): Mean CRB for signals with (c)  $\text{SNR}_1 = 10 \text{ dB}$  and  $\text{SNR}_2 = 5 \text{ dB}$  and (d)  $\text{SNR}_2 = 15 \text{ dB}$ .



constraint that the  $N - T$  unknown signals are CM, i.e.,  $|s_i(t)| = 1$ ,  $i = 1, \dots, k$ ,  $t = T + 1, \dots, N$ . We can also apply the constraint of known  $H$  matrix, which provides a basis for evaluating the effectiveness of the  $T$  training symbols for estimation of  $H$ .

#### 4.2. Example

Consider an example with  $k = 2$  transmit antennas,  $l = 2$  receive antennas, independent Rayleigh fading, and  $N = 50$  time samples with  $T = 2$  training symbols. The fading is assumed to be constant over the block of  $N$  symbol times. The SNR per receive antenna is varied over the range 0 to 20 dB, and the constrained CRBs are averaged over 500 independent fading matrices  $H$  for each SNR value. The signals are 8-PSK, and the transmitted signals satisfy the space-time code constraint (19). For each realization of  $H$ , we compute CRBs on the signal phases  $\angle s_{T+1}, \dots, \angle s_N$  subject to various constraints, and these CRBs are averaged to obtain mean CRBs for the realization.

Figure 2 contains constrained CRBs on signal phase estimation for various constraints. The space-time code structure (19) is present in the transmitted signals, but it is only enforced in the constraints labeled 'S-T CODE'. When the space-time code is not applied, the CRB corresponds to independent estimation of the transmitted sequences  $s_1(T+1), \dots, s_1(N)$  and  $s_2(T+1), \dots, s_2(N)$ , so diversity gain is impossible. We make the following observations from Figure 2.

- Comparing 'KNOWN H' with 'KNOWN H & S-T CODE' shows a potential diversity gain of approximately 10 dB in SNR provided by the space-time code when  $H$  is known exactly.
- Comparing 'SEMI-BLIND & S-T CODE' with 'KNOWN H & S-T CODE' shows that  $T = 2$  training symbols for estimation of  $H$  costs approximately 3 dB in SNR compared with exact knowledge of  $H$ .
- The 'SEMI-BLIND & CM & S-T CODE' curve shows that exploiting CM in addition to the training and space-time code potentially yields about 1.5 dB gain in SNR.
- For the cases in which the space-time code constraint is not exploited, the 'SEMI-BLIND & CM' constraint provides approximately 2 dB gain compared with 'KNOWN H', which does not exploit the CM property.

Note that the constrained CRBs on  $\angle s_t$  pertain to estimation of the signals, while the primary quantity of interest in digital communication is probability of detection error. Smaller CRBs suggest the potential for reduced probability of detection error in practical receivers.

#### 5. REFERENCES

[1] B.M. Sadler, R.J. Kozick, T. Moore, "Bounds on constant modulus and semi-blind array processing," Proc. CISS'2000, March 2000.

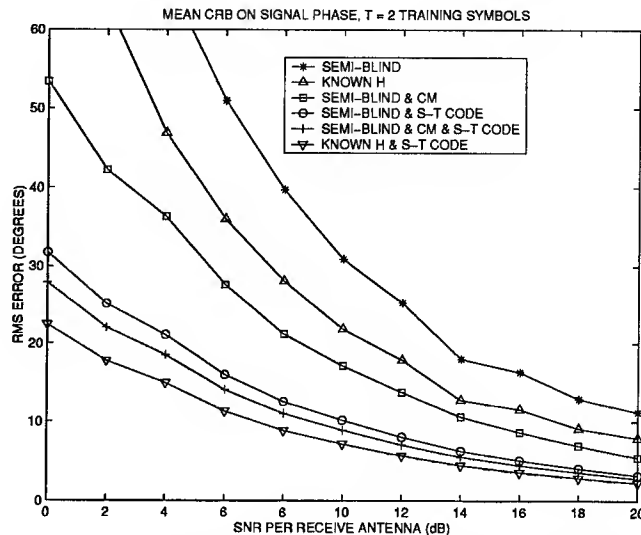


Figure 2: CRBs for signal phase estimation in space-time coding scenario with  $k = 2$  transmitters,  $l = 2$  receivers,  $N = 50$  time samples, and independent Rayleigh fading with various constraints.

[2] B.M. Sadler, R.J. Kozick, T. Moore, "Performance bounds on bearing and symbol estimation for communication signals with side information," Proc. ICASSP 2000, June 2000.

[3] P. Stoica, B. C. Ng, "On the Cramér-Rao bound under parametric constraints," IEEE Sig. Proc. Letters, vol. 5, no. 7, pp. 177-179, July 1998.

[4] J.F. Cardoso and A. Souloumiac, "Blind beamforming for non-Gaussian signals," IEE Proc.F, Vol. 140, No. 6, pp. 362-370, Dec. 1993.

[5] P. Comon, "Independent component analysis, A new concept?," Signal Processing, vol. 36, pp. 287-314, 1994.

[6] J. Sheinvald, "On blind beamforming for multiple non-Gaussian signals and the constant-modulus algorithm," IEEE Trans. Signal Processing, vol. 46, no. 7, pp. 1878-1885, July 1998.

[7] A.-J. van der Veen, A. Paulraj, "An analytical constant modulus algorithm," IEEE Trans. Signal Processing, vol. 44, no. 5, pp. 1136-1155, May 1996.

[8] A. F. Naguib, N. Seshadri, A. R. Calderbank, "Increasing data rate over wireless channels," IEEE Sig. Proc. Mag., May 2000.

[9] S. M. Alamouti, "A simple transmit diversity technique for wireless communications," IEEE J. on Selected Areas in Comm., Oct. 1998.

# ARRAY PROCESSING IN THE PRESENCE OF UNKNOWN NONUNIFORM SENSOR NOISE: A MAXIMUM LIKELIHOOD DIRECTION FINDING ALGORITHM AND CRAMÉR-RAO BOUNDS

Marius Pesavento

Alex B. Gershman

Department of ECE, McMaster University  
Hamilton, Ontario, L8S 4K1 Canada  
gershman@ieee.org

## ABSTRACT

We address the problem of estimating Directions Of Arrival (DOA's) of multiple sources observed on the background of nonuniform white noise with an arbitrary unknown diagonal covariance matrix. A new deterministic Maximum Likelihood (ML) DOA estimator is derived. Its implementation is based on an iterative procedure which includes stepwise concentration of the Log-Likelihood (LL) function with respect to the signal and noise nuisance parameters and requires only a few iterations to converge.

New closed-form expressions for the deterministic and stochastic direction estimation Cramér-Rao bounds (CRB's) are derived for the considered nonuniform model. Our expressions can be viewed as an extension of the well-known results by Stoica and Nehorai, and Weiss and Friedlander to a more general noise model than the commonly used uniform one. Simulation and experimental (seismic data processing) results illustrate the performance of the estimator and validate our theoretical analysis.

## 1. INTRODUCTION

ML DOA estimation techniques are known to have excellent asymptotic and threshold performances [1], [2]. The key assumption used for the derivation of both the deterministic and stochastic ML estimators is the so-called *uniform* white noise assumption [1]. According to it, sensor noises are presumed to form a zero-mean Gaussian process with the covariance matrix  $\sigma^2 \mathbf{I}$ , where  $\sigma^2$  is the unknown noise variance, and  $\mathbf{I}$  is the identity matrix. This simple assumption enables to concentrate the resulting LL function with respect to both signal waveform and noise nuisance parameters, and, therefore, reduce the dimension of the parameter space and the associated computational burden [1].

Apparently, the uniform noise assumption may be unrealistic in certain applications [3]-[6], where the noise environment remains unknown or changes slowly with time. In the general case, the sensor noise should be considered as an unknown colored (i.e. spatially dependent) process. Recently, several advanced ML techniques have been proposed which exploit the ideas of colored noise modeling [6]-[8].

In some practical applications (for example, when the so-called *sparse* arrays are used), the general colored noise

assumption can be simplified by assuming the sensor noise to be spatially white [4], [5]. In this case, the noise spatial covariance structure still can be represented by a diagonal matrix but the sensor noise variances are no longer identical one to another. Such a noise model becomes relevant in situations with hardware nonidealities in receiving channels [9] as well as for sparse arrays with prevailing external noise (for example, reverberation noise in sonar or external seismic noise) [4], [5].

It is important to stress that if sensor noise is a spatially nonuniform white process, neither the conventional "uniform" ML methods [1]-[2], nor the colored noise modeling ML techniques [6]-[8] may be expected to give satisfactory results, because the former methods will mismodel the noise, whereas the latter techniques will ignore important *prior* knowledge that the noise process is spatially white. This appears to be a strong motivation to develop direction finding techniques for the nonuniform white noise case. Moreover, the majority of the ML colored noise modeling based approaches developed so far are unable to concentrate the LL function with respect to the noise parameters [7]. As a result, such techniques may be computationally demanding. The use of the nonuniform white noise model can be expected to overcome this drawback by means of obtaining "concentrated" solutions to the ML estimation problem.

The motivation given shows that the nonuniform white noise case can be viewed as a practically important generalization of the simpler uniform model. In the present paper, we derive a new iterative deterministic ML estimator, which concentrates the LL function with respect to both signal and noise nuisance parameters. Unlike the analytic concentration used in the conventional "uniform" ML estimators, the concentration of the LL function in the nonuniform noise case will be performed in a numerical (iterative) manner, with only a few iterations necessary for convergence.

Furthermore, we derive closed-form expressions for the deterministic and stochastic direction estimation CRB's for the considered nonuniform white noise case. These expressions can be viewed as a natural extension of the well-known results reported in [1]-[2] and [10] for the uniform noise model. The estimation performance of the proposed ML technique is compared to the derived CRB's and the performance of the deterministic uniform ML estimator [1] via

This work was supported by the Natural Sciences and Engineering Research Council (NSERC) of Canada.

computer simulations. Moreover, we test both the uniform and nonuniform ML techniques using experimental seismic data recorded by the GERESS array (Germany). Our simulations and the results of real data processing demonstrate essential performance improvements achieved by means of the proposed nonuniform ML estimator relative to the conventional uniform ML algorithm. Additionally, the experimental results provide a solid verification of the practical relevance of the considered nonuniform noise model.

## 2. SIGNAL MODEL

Let an array of  $n$  sensors receive  $q$  ( $q < n$ ) narrowband signals impinging from the sources with unknown DOA's  $\theta_1, \dots, \theta_q$ . The  $i$ th snapshot vector of sensor array outputs can be modeled as [1]-[3]

$$\mathbf{x}(i) = \mathbf{A}(\theta)\mathbf{s}(i) + \mathbf{n}(i), \quad i = 1, \dots, N \quad (1)$$

where  $\mathbf{A}(\theta) = [\mathbf{a}(\theta_1), \dots, \mathbf{a}(\theta_q)]$  is the  $n \times q$  matrix composed from the signal direction vectors  $\mathbf{a}(\theta_i)$  ( $i = 1, \dots, q$ ),  $\theta = [\theta_1, \dots, \theta_q]^T$  is the  $q \times 1$  vector of the unknown signal DOA's,  $\mathbf{s}(i)$  is the  $q \times 1$  vector of the source waveforms,  $\mathbf{n}(i)$  is the  $n \times 1$  vector of white sensor noise,  $N$  is the number of snapshots, and  $(\cdot)^T$  stands for the transpose. In a more compact notation, (1) can be rewritten as

$$\mathbf{X} = \mathbf{A}(\theta)\mathbf{S} + \mathbf{N} \quad (2)$$

where  $\mathbf{X} = [\mathbf{x}(1), \dots, \mathbf{x}(N)]$  is the  $n \times N$  array data matrix,  $\mathbf{S} = [\mathbf{s}(1), \dots, \mathbf{s}(N)]$  is the  $q \times N$  source waveform matrix, and  $\mathbf{N} = [\mathbf{n}(1), \dots, \mathbf{n}(N)]$  is the  $n \times N$  sensor noise matrix. The sensor noise is assumed to be a zero-mean spatially and temporally white Gaussian process with the unknown diagonal covariance matrix

$$\mathbf{Q} = \mathbf{E}\{\mathbf{n}(t)\mathbf{n}^H(t)\} = \text{diag}\{\sigma_1^2, \sigma_2^2, \dots, \sigma_n^2\} \quad (3)$$

In what follows, the signal waveforms will be assumed to be either deterministic unknown processes [1], or random zero-mean Gaussian processes [2]. In particular, the signal snapshots are assumed to satisfy the following models

$$\mathbf{x}(i) \sim \mathcal{N}(\mathbf{A}\mathbf{s}(i), \mathbf{Q}) \quad (4)$$

$$\mathbf{x}(i) \sim \mathcal{N}(\mathbf{0}, \mathbf{R}) \quad (5)$$

in the deterministic and stochastic case, respectively. Here,

$$\mathbf{R} = \mathbf{E}\{\mathbf{x}(i)\mathbf{x}^H(i)\} = \mathbf{A}\mathbf{P}\mathbf{A}^H + \mathbf{Q} \quad (6)$$

is the array covariance matrix,  $\mathbf{P} = \mathbf{E}\{\mathbf{s}(i)\mathbf{s}^H(i)\}$  is the source waveform covariance matrix,  $\mathcal{N}$  denotes the complex Gaussian distribution, and  $(\cdot)^H$  stands for the Hermitian transpose.

## 3. MAXIMUM LIKELIHOOD ESTIMATION

Under the assumption that the signal waveforms are deterministic unknown sequences, the LL function for the model considered is given by [11]

$$L(\Psi) = -N \sum_{k=1}^n \log \sigma_k^2 - \sum_{i=1}^N \|\tilde{\mathbf{x}}(i) - \tilde{\mathbf{A}}(\theta)\mathbf{s}(i)\|^2 \quad (7)$$

where  $\Psi = [\theta^T, \sigma^T, \mathbf{s}^T(1), \dots, \mathbf{s}^T(N)]^T$  is the vector of unknown signal and noise parameters,  $\sigma = [\sigma_1^2, \dots, \sigma_n^2]^T$ ,  $\tilde{\mathbf{x}}(i) = \mathbf{Q}^{-1/2}\mathbf{x}(i)$ , and  $\tilde{\mathbf{A}}(\theta) = \mathbf{Q}^{-1/2}\mathbf{A}(\theta)$ .

Introduce the  $n \times N$  matrix

$$\mathbf{G} = \mathbf{X} - \mathbf{A}(\theta)\mathbf{S} = [\mathbf{c}_1, \dots, \mathbf{c}_N] = [\mathbf{r}_1, \dots, \mathbf{r}_n]^T \quad (8)$$

where  $\mathbf{c}_i$  and  $\mathbf{r}_l$  are the  $n \times 1$  and  $N \times 1$  vectors corresponding to the  $i$ th column and the  $l$ th row of the matrix  $\mathbf{G}$ , respectively. With these notations, from (7) it follows that

$$\frac{\partial L(\Psi)}{\partial \sigma_k^2} = \mathbf{e}_k^T \left( \sum_{i=1}^N \mathbf{c}_i \mathbf{c}_i^H \mathbf{Q}^{-1} - N \mathbf{I} \right) \mathbf{Q}^{-1} \mathbf{e}_k \quad (9)$$

where  $\mathbf{e}_k$  is the vector containing one in the  $k$ th position and zeros elsewhere.

From (3) and (9), we obtain that if the other parameters are fixed, the ML estimate of the diagonal noise covariance matrix is given by

$$\hat{\mathbf{Q}} = \frac{1}{N} \text{diag}\{\mathbf{r}_1^H \mathbf{r}_1, \mathbf{r}_2^H \mathbf{r}_2, \dots, \mathbf{r}_n^H \mathbf{r}_n\} \quad (10)$$

Here, we exploit the following obvious property  $[\mathbf{C}]_{k,k} = \mathbf{r}_k^H \mathbf{r}_k$  of the matrix

$$\mathbf{C} = \sum_{i=1}^N \mathbf{c}_i \mathbf{c}_i^H \quad (11)$$

Inserting (10) into (7), we have

$$L(\theta, \mathbf{S}) = -N \sum_{k=1}^n \log \left\{ \frac{1}{N} \mathbf{r}_k^H \mathbf{r}_k \right\} - \sum_{i=1}^N \mathbf{c}_i^H \hat{\mathbf{Q}}^{-1} \mathbf{c}_i \quad (12)$$

Using (10)-(11) and the properties of the trace operator, we obtain that

$$\begin{aligned} \sum_{k=1}^N \mathbf{c}_k^H \hat{\mathbf{Q}}^{-1} \mathbf{c}_k &= \sum_{k=1}^N \text{trace} \left\{ \hat{\mathbf{Q}}^{-1} \mathbf{c}_k \mathbf{c}_k^H \right\} \\ &= \text{trace} \left\{ \hat{\mathbf{Q}}^{-1} \mathbf{C} \right\} = nN \end{aligned} \quad (13)$$

Hence, after omitting the constant term (13), the LL function (12) can be further simplified to

$$L(\theta, \mathbf{S}) = -N \sum_{k=1}^n \log \left\{ \frac{1}{N} \mathbf{r}_k^H \mathbf{r}_k \right\} \quad (14)$$

At the same time, from (7) we obtain in a standard way that if the remaining parameters are fixed, the ML estimate of the matrix  $\mathbf{S}$  is given by

$$\hat{\mathbf{S}} = \left( \tilde{\mathbf{A}}^H(\theta) \tilde{\mathbf{A}}(\theta) \right)^{-1} \tilde{\mathbf{A}}^H(\theta) \tilde{\mathbf{X}} \quad (15)$$

where  $\tilde{\mathbf{X}} = \mathbf{Q}^{-1/2} \mathbf{X}$  is the  $n \times N$  transformed data matrix. Note that the estimate (15) depends on  $\mathbf{Q}$ , and, in turn, the estimate of  $\mathbf{Q}$  in (10) depends on  $\mathbf{S}$ . Therefore, it appears to be impossible to obtain any closed form expression of the LL function concentrated with respect to the full set

of the signal and noise nuisance parameters. To avoid this difficulty, we introduce the idea of *stepwise concentration*, which was also exploited in [3] in an implicit form. The essence of this idea is to concentrate the LL function in an iterative manner.

Omitting the constant factor  $-N$  in (14) and inserting (15) into this equation, we obtain the following alternative expressions for the negative LL function

$$\begin{aligned}\mathcal{L}(\theta) &= \sum_{k=1}^n \log \left\{ \frac{1}{N} \mathbf{r}_k^H \mathbf{r}_k \right\} \\ &= \text{trace} \log \left\{ \frac{1}{N} \mathbf{G} \mathbf{G}^H \right\} \\ &= \text{trace} \log \left\{ \frac{1}{N} \mathbf{P}_{\tilde{\mathbf{A}}}^\perp(\theta) \tilde{\mathbf{X}} \tilde{\mathbf{X}}^H \mathbf{P}_{\tilde{\mathbf{A}}}^\perp(\theta) \right\} \\ &= \text{trace} \log \left\{ \mathbf{P}_{\tilde{\mathbf{A}}}^\perp(\theta) \hat{\tilde{\mathbf{R}}} \right\}\end{aligned}\quad (16)$$

where  $\mathbf{P}_{\tilde{\mathbf{A}}}^\perp(\theta) = \tilde{\mathbf{A}}(\theta) \left( \tilde{\mathbf{A}}^H(\theta) \tilde{\mathbf{A}}(\theta) \right)^{-1} \tilde{\mathbf{A}}^H(\theta)$  and  $\mathbf{P}_{\tilde{\mathbf{A}}}^\perp(\theta) = \mathbf{I} - \mathbf{P}_{\tilde{\mathbf{A}}}(\theta)$  are the projection matrices. Here,

$$\hat{\tilde{\mathbf{R}}} = \frac{1}{N} \tilde{\mathbf{X}} \tilde{\mathbf{X}}^H \quad (17)$$

is the  $n \times n$  sample covariance matrix of the transformed data.

It is important to stress that in the particular uniform noise case ( $\mathbf{Q} = \sigma^2 \mathbf{I}$ ), the function (16) can be simplified to

$$\mathcal{L}(\theta) = \text{trace} \log \left\{ \mathbf{P}_{\tilde{\mathbf{A}}}^\perp(\theta) \hat{\tilde{\mathbf{R}}} \right\} \quad (18)$$

where

$$\hat{\tilde{\mathbf{R}}} = \frac{1}{N} \mathbf{X} \mathbf{X}^H \quad (19)$$

is the sample covariance matrix of the original data (1). Interestingly, this function is not equivalent to the conventional negative LL function [1]

$$\mathcal{L}(\theta) = \text{trace} \left\{ \mathbf{P}_{\tilde{\mathbf{A}}}^\perp(\theta) \hat{\tilde{\mathbf{R}}} \right\} \quad (20)$$

derived under the uniform noise assumption. The explanation of this fact lies on the basis of the observation that the ML estimators (16) and (20) use very different types of *a priori* information on the structure of the noise covariance matrix.

Another important observation is that unlike (20), the function (16) does not enable simultaneous concentration with respect to both signal and noise nuisance parameters. This fact can be explained by inspecting the structure of (16). According to this equation, the estimate of the signal DOA vector  $\theta$  depends on the estimate (10) of the matrix  $\mathbf{Q}$ , which, in turn, is dependent of the estimate of  $\theta$ . To overcome this problem, instead of the analytic concentration approach used for the derivation of the uniform ML estimator, we propose the so-called *stepwise numerical concentration*, which is given by the following iterative procedure:

- **Step 1.** Set  $\hat{\mathbf{Q}} = \mathbf{I}$ .
- **Step 2.** Find the estimate of  $\theta$  as  $\hat{\theta} = \text{argmin}_{\theta} \{ \mathcal{L}(\theta) \}$  where the negative LL function  $\mathcal{L}(\theta)$  is defined by (16).
- **Step 3.** Using the so-obtained  $\hat{\theta}$ , compute  $\hat{\mathbf{S}}$  from (15). Find the refined estimate of  $\mathbf{Q}$  from (10) using (8) and the previously obtained (fixed)  $\hat{\mathbf{S}}$  and  $\hat{\theta}$ . Repeat steps 2 and 3 a few times to obtain the final estimate of  $\theta$ .

In step 1, the algorithm is initialized using the uniform noise assumption. Under this assumption, the estimate of  $\mathbf{Q}$  should be written as  $\hat{\mathbf{Q}} = \hat{\sigma}^2 \mathbf{I}$ , where  $\hat{\sigma}^2$  is some estimate of the noise variance  $\sigma^2$ . However, from the structure of the negative LL function (16) it follows that the minimizer of this function does not depend on the value of  $\hat{\sigma}^2$ . Therefore, without loss of generality in step 1 we can set  $\hat{\sigma}^2 = 1$ .

#### 4. CRAMÉR-RAO BOUNDS

The following two theorems present closed-form expressions for the deterministic and stochastic CRB's under the nonuniform noise assumption.

*Theorem 1:* The  $q \times q$  deterministic CRB matrix for the signal DOA's is given by:

$$\text{CRB}_{\text{DET}} \theta \theta = \frac{1}{2N} \left\{ \text{Re} \left[ \left( \tilde{\mathbf{D}}^H \mathbf{P}_{\tilde{\mathbf{A}}}^\perp \tilde{\mathbf{D}} \right) \odot \hat{\mathbf{P}}^T \right] \right\}^{-1} \quad (21)$$

where  $\tilde{\mathbf{A}} = \mathbf{Q}^{-1/2} \mathbf{A}$ ,  $\tilde{\mathbf{D}} = \mathbf{Q}^{-1/2} \mathbf{D}$ ,  $\hat{\mathbf{P}} = \frac{1}{N} \sum_{i=1}^N \mathbf{s}(i) \mathbf{s}(i)^H$ ,  $\odot$  stands for the Schur-Hadamard matrix product, and

$$\mathbf{D} = \left[ \left. \frac{d\mathbf{a}(\theta)}{d\theta} \right|_{\theta=\theta_1}, \left. \frac{d\mathbf{a}(\theta)}{d\theta} \right|_{\theta=\theta_2}, \dots, \left. \frac{d\mathbf{a}(\theta)}{d\theta} \right|_{\theta=\theta_q} \right] \quad (22)$$

*Proof:* See [11].

*Theorem 2:* The  $q \times q$  stochastic CRB matrix for the signal DOA's is given by:

$$\begin{aligned}\text{CRB}_{\text{STO}} \theta \theta &= \frac{1}{N} \left\{ 2 \text{Re} \left[ \left( \mathbf{P}_{\tilde{\mathbf{A}}} \tilde{\mathbf{R}}^{-1} \tilde{\mathbf{A}} \mathbf{P} \right) \right. \right. \\ &\quad \left. \left. \odot \left( \tilde{\mathbf{D}}^H \mathbf{P}_{\tilde{\mathbf{A}}}^\perp \tilde{\mathbf{R}}^{-1} \tilde{\mathbf{D}} \right)^T \right] - \mathbf{M} \mathbf{T} \mathbf{M}^T \right\}^{-1}\end{aligned}\quad (23)$$

where  $\tilde{\mathbf{R}} = \mathbf{Q}^{-1/2} \mathbf{R} \mathbf{Q}^{-1/2}$  and the real matrices

$$\mathbf{M} = 2 \text{Re} \left\{ \left( \tilde{\mathbf{R}}^{-1} \tilde{\mathbf{A}} \mathbf{P} \right)^T \odot \left( \tilde{\mathbf{D}}^H \mathbf{P}_{\tilde{\mathbf{A}}}^\perp \right) \right\}, \quad (24)$$

$$\begin{aligned}\mathbf{T} &= \left\{ \left( \tilde{\mathbf{R}}^{-1} \right)^* \odot \tilde{\mathbf{R}}^{-1} \right. \\ &\quad \left. - \left( \mathbf{P}_{\tilde{\mathbf{A}}}^\perp \tilde{\mathbf{R}}^{-1} \right)^* \odot \left( \mathbf{P}_{\tilde{\mathbf{A}}}^\perp \tilde{\mathbf{R}}^{-1} \right) \right\}^{-1}\end{aligned}\quad (25)$$

*Proof:* See [11].

It is interesting to compare the derived expressions with the deterministic and stochastic CRB's in the uniform noise

case. The latter two bounds are given by [1], [2], [10]

$$\text{CRB}_{\text{DET}} \theta\theta = \frac{\sigma^2}{2N} \left\{ \text{Re} \left[ (D^H P_A^\perp D) \odot \hat{P}^T \right] \right\}^{-1} \quad (26)$$

$$\text{CRB}_{\text{STO}} \theta\theta = \frac{\sigma^2}{2N} \left\{ \text{Re} \left[ (P A^H R^{-1} A P) \odot (D^H P_A^\perp R^{-1} D)^T \right] \right\}^{-1} \quad (27)$$

respectively.

The comparison of (21) and (26) shows that the nonuniform deterministic bound (21) corresponds to the uniform CRB (26), with the only difference that the nonuniform CRB uses the transformed array manifold  $\tilde{A}$  instead of the original manifold  $A$ . This transformation can be viewed as a sort of *preequalization* of sensor noise<sup>1</sup>. To explain the effect of noise preequalization, let us consider the case when some part of array sensors suffers from intensive noises, whereas another part of sensors remains relatively "noiseless". According to the above-mentioned manifold transformation, the contribution of the noisy sensors to the CRB (21) will be negligible because of relatively low weights assigned to these sensors. This corresponds to our natural expectation that the optimal (ML) algorithm derived for the nonuniform model should be insensitive to the presence of such noisy sensors. Such a robustness property is achieved by means of blocking the outputs of corresponding (noisy) array channels and exploiting only noiseless sensors. From this point of view, the manifold transformation matrix  $Q^{-1/2}$  can be identified as a sort of *blocking matrix*.

As it can be seen from the comparison of (23) and (27), in the stochastic case the relationship between the uniform and nonuniform bounds becomes more complicated than in the deterministic case. In particular, this relationship cannot be described solely in terms of the manifold transformation  $Q^{-1/2}$ . We observe that the bound (23) contains an additional term  $-MTM^T$  which does not appear in (27). In the general case, we obtain that

$$\text{Nonuniform CRB}_{\text{DET}} \theta\theta \Big|_{Q=\sigma^2 I} = \text{Uniform CRB}_{\text{DET}} \theta\theta$$

$$\text{Nonuniform CRB}_{\text{STO}} \theta\theta \Big|_{Q=\sigma^2 I} \geq \text{Uniform CRB}_{\text{STO}} \theta\theta$$

The proof of the last equation is given in [11].

Assume that there is only one signal source ( $q = 1$ ). In this case, we have that  $\tilde{A} = \tilde{a}$  and  $\tilde{D} = \tilde{d}$ , where  $\tilde{a} = Q^{-1/2}a$ . Therefore, the array covariance matrix (6) can be rewritten as  $R = p a a^H + Q$ , where  $p = E\{|s(i)|^2\}$  is the signal variance. It is easy to show that in this case the bounds (21) and (23) can be simplified to [5]

$$\text{CRB}_{\text{DET}} \theta\theta = \frac{a^H Q^{-1} a}{2N \hat{p} [a^H Q^{-1} a a^H B^2 Q^{-1} a - (a^H B Q^{-1} a)^2]}$$

$$\text{CRB}_{\text{STO}} \theta\theta = \frac{1 + p a^H Q^{-1} a}{2N p^2 [a^H Q^{-1} a a^H B^2 Q^{-1} a - (a^H B Q^{-1} a)^2]}$$

<sup>1</sup>Usually, the term *preequalization* is used but this is somewhat confusing to use it here because sensor noise has been originally assumed to be spatially white.

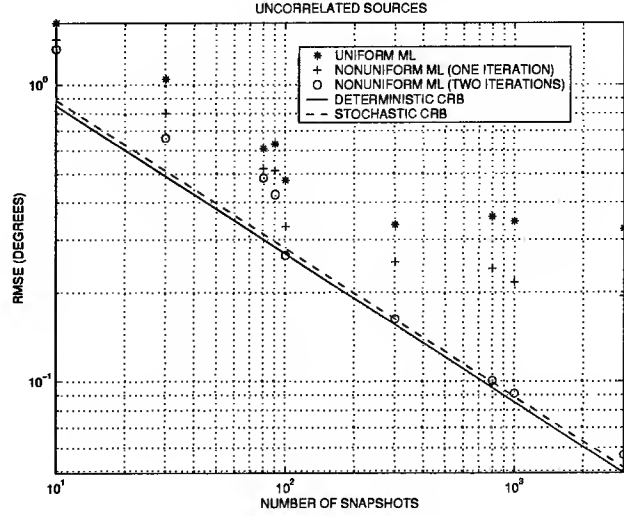


Figure 1: Comparison of the DOA estimation RMSE's and CRB's. First example.

where  $B = \text{diag}\{(\omega/c)d_1 \cos \theta_1, (\omega/c)d_2 \cos \theta_1, \dots, (\omega/c)d_n \cos \theta_1\}$ ,  $\hat{p} = \frac{1}{N} \sum_{i=1}^N |s(i)|^2$ ,  $d_k$  is the coordinate of the  $k$ -th sensor,  $\omega$  is the central frequency, and  $c$  is the propagation speed.

Assuming that the array has omnidirectional sensors, the number of snapshots is high ( $\hat{p} \simeq p$ ), and defining the SNR as [5]  $\text{SNR} = (p/n) a^H Q^{-1} a = (p/n) \sum_{i=1}^n 1/\sigma_i^2$ , we obtain the following explicit relationship between the stochastic and deterministic single-source bounds:

$$\text{CRB}_{\text{STO}} \theta\theta \simeq \left(1 + \frac{1}{n \text{SNR}}\right) \text{CRB}_{\text{DET}} \theta\theta \quad (28)$$

Hence, in the large sample case the difference between the two bounds becomes small when the source is powerful enough, so that  $n \text{SNR} \gg 1$ .

## 5. SIMULATIONS

We assumed a ULA of ten sensors spaced half a wavelength apart, and two equally powered sources with the DOA's  $\theta_1 = 7^\circ$  and  $\theta_2 = 13^\circ$ . The nonuniform noise was assumed to have the following covariance matrix:  $Q = \text{diag}\{10.0, 2.0, 1.5, 0.5, 8.0, 0.7, 1.1, 3.0, 6.0, 3.0\}$ . In all our examples, the experimental DOA estimation Root-Mean-Square Errors (RMSE's) of the conventional uniform and the proposed nonuniform ML methods have been compared to the nonuniform CRB's (21) and (23).

In the first example, we assume two uncorrelated sources with the SNR = 10 dB. Fig. 1 displays the results versus the number of snapshots. In the second example, two correlated sources are assumed, with the correlation coefficient equal to 0.9. The SNR = 15 dB is taken and the results are plotted in Fig. 2 versus the number of snapshots.

From Figs. 1-2, we observe that uniform ML performs poorly in the nonuniform noise case. As expected, the proposed nonuniform technique provides essential performance improvements. In particular, it attains the stochastic CRB

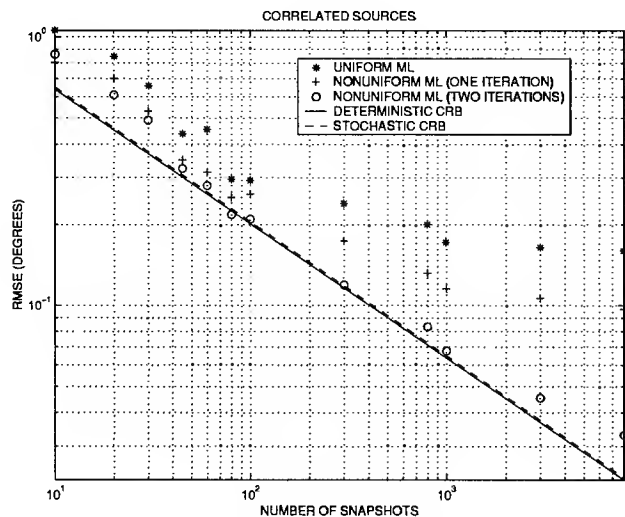


Figure 2: Comparison of the DOA estimation RMSE's and CRB's. Second example.

(23) even at small sample sizes. Since two iterations are enough to guarantee the convergence, the computational cost of our technique is comparable to that of conventional ML.

## 6. EXPERIMENTAL RESULTS

To validate the practical relevance of the nonuniform noise model, real seismic data were used. These data were collected by GERESS array (Germany). The data record of the regional seismic event at an azimuth of  $\theta = 121.8^\circ$  was analyzed (see [12] for details). Note that the azimuth value of this event was known in advance with a high precision. Estimating this parameter using the methods tested, we were able to compare their experimental performances.

The conventional and proposed ML methods have been applied to azimuth-velocity (2D) estimation at the following four frequencies:  $f_1 = 0.9375$  Hz,  $f_2 = 1.25$  Hz,  $f_3 = 1.5625$  Hz, and  $f_4 = 1.875$  Hz.

The experimental azimuth estimates have been used to compute the experimental RMSE's shown in Fig. 3. From this figure, it is clearly seen that nonuniform ML has noticeably better experimental performance than the uniform ML technique. These results provide a solid verification of relevance of the developed nonuniform noise model in practical applications.

## REFERENCES

- [1] P. Stoica and A. Nehorai, "MUSIC, maximum likelihood and Cramér-Rao bound," *IEEE Trans. ASSP*, **37**, pp. 720-741, May 1989.
- [2] P. Stoica and A. Nehorai, "Performance study of conditional and unconditional direction-of-arrival estimation," *IEEE Trans. ASSP*, **38**, pp. 1783-1795, Oct. 1990.

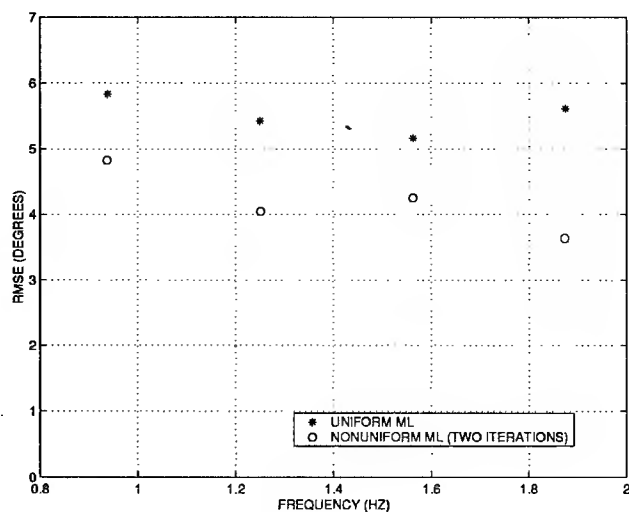


Figure 3: Comparison of the DOA estimation RMSE's. Real seismic array data.

- [3] J.F. Böhme and D. Kraus, "On least squares methods for direction of arrival estimation in the presence of unknown noise fields", *ICASSP'88*, NY, pp. 2833-2836, Apr. 1988.
- [4] A.B. Gershman, A.L. Matveyev, and J.F. Böhme, "Maximum likelihood estimation of signal power in sensor array in the presence of unknown noise field," *IEE Proc. RSN*, **F-142**, pp. 218-224, Oct. 1995.
- [5] A.L. Matveyev, A.B. Gershman, and J.F. Böhme, "On the direction estimation Cramer-Rao bounds in the presence of uncorrelated unknown noise," *Circ., Syst., Signal Processing*, **18**, pp. 479-487, 1999.
- [6] J. LeCadre, "Parametric methods for spatial signal processing in the presence of unknown colored noise fields," *IEEE Trans. ASSP*, **37**, pp. 965-983, July 1989.
- [7] B. Friedlander and A.J. Weiss, "Direction finding using noise covariance modeling," *IEEE Trans. SP*, **43**, pp. 1557-1567, July 1995.
- [8] P. Stoica, M. Viberg, K.M. Wong, and Q. Wu, "Maximum-likelihood bearing estimation with partly calibrated arrays in spatially correlated noise field," *IEEE Trans. SP*, **44**, pp. 888-899, Apr. 1996.
- [9] U. Nickel, "On the influence of channel errors on array signal processing methods", *Int. J. Electron. and Comm.*, **47**, pp. 209-219, 1993.
- [10] A.J. Weiss and B. Friedlander, "On the Cramér-Rao bound for direction finding of correlated sources", *IEEE Trans. SP*, **41**, pp. 495-499, Jan. 1993.
- [11] M. Pesavento and A.B. Gershman, "Maximum-likelihood direction of arrival estimation in the presence of unknown nonuniform noise", submitted.
- [12] D.V. Sidorovich and A.B. Gershman, "2-D wideband interpolated root-MUSIC applied to measured seismic data", *IEEE Trans. SP*, **46**, pp. 2263-2267, Aug. 1998.

# MATCHED SYMMETRICAL SUBSPACE DETECTOR

Victor S. Golikov, Francisco C. Pareja

Ciencia y Tecnologia del Mayab, A. C.

Calle 12, No. 199, dep.5, entre 19 y 21, Col. Garcia Gineres, C.P. 97070, Merida, Yucatan, Mexico

## ABSTRACT

The optimal detection/estimation algorithms require large computing expenditures in the radar, sonar and etc. The paper presents the new Uniformly Most Powerful Test for matched detecting of the symmetrical signal subspace. The general (logical) shift operators group is used for describing of the symmetry. This algorithm may be used to reduce the complexity of matched detector for unknown signal subspace and for a signal processing in real time. The reduction brings appreciable hardware gains and a small performance penalties in some radar systems. The signal subspace model for moving-target indication in radar is considered. We used the new approach for creation of the sub-optimal detector with minimal computing expenditures.

## 1. INTRODUCTION

In signal detection problems, we assume that each measurement is a sum of a signal component and a noise component:

$$x_n = \mu s_n + \sigma w_n; \quad n=0, 1, \dots, N-1.$$

The measurements are organized into a N-dimensional measurement vector

$$\mathbf{x} = \mu \mathbf{s} + \sigma \mathbf{w}; \quad (1)$$

where vector  $\mu \mathbf{s}$  contains samples of the signal to be detected and the vector  $\sigma \mathbf{w}$  contains samples of the added noises. We assume that the noise vector  $\mathbf{w}$  is draw from a multivariate normal distribution  $\mathbf{w} \sim N[0, \mathbf{I}]$ . This means that the measurement  $\mathbf{x}$  is drawn from a multivariate normal distribution  $\mathbf{x} \sim N[\mu \mathbf{s}, \sigma^2 \mathbf{I}]$ . In some systems it sometimes happens that the signal  $\mathbf{s}$  in the measurement model  $\mathbf{x} \sim N[\mu \mathbf{s}, \sigma^2 \mathbf{I}]$  is a linear combination of modes or basis vectors, in which case it may be represented as

$$\mathbf{s} = \sum_{n=0}^{N-1} \theta_n \mathbf{h}_n = \mathbf{H} \boldsymbol{\theta}. \quad \text{Here } \mathbf{H} \text{ is a known } N \times N \text{ matrix with}$$

columns  $\mathbf{h}_n$  and  $\boldsymbol{\theta}$  is a unknown  $N \times 1$  vector with elements  $\theta_n$ :

$$\mathbf{s} = [\mathbf{h}_0 \ \mathbf{h}_1 \ \dots \ \mathbf{h}_{N-1}] \begin{bmatrix} \theta_0 \\ \vdots \\ \theta_{N-1} \end{bmatrix}. \quad (2)$$

Let the mode matrix  $\mathbf{H}$  is known but the mode weights are unknown. In this case, the signal is known to lie in the linear subspace  $\langle \mathbf{H} \rangle$  spanned by the columns of  $\mathbf{H}$ , but its exact location is unknown because  $\boldsymbol{\theta}$  is unknown. We would like to test  $H_0: \mu=0$  versus  $H_1: \mu>0$  when  $\mathbf{x}$  is distributed as  $N[\mu \mathbf{H} \boldsymbol{\theta}, \sigma^2 \mathbf{I}]$  and  $\boldsymbol{\theta}$  is unknown. It is known [1] that the statistic  $\chi^2 = \mathbf{x}^T \mathbf{P}_H \mathbf{x}$  (3) is a maximal invariant to the group of transformations that adds a bias from the orthogonal subspace  $\langle \mathbf{A} \rangle$  and rotates in the

subspace  $\langle \mathbf{H} \rangle$ . Here  $\mathbf{P}_H$  is the projection  $\mathbf{x}$  on subspace  $\langle \mathbf{H} \rangle$ :

$$\mathbf{P}_H = \mathbf{H}(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T. \quad (4)$$

The statistic  $\chi^2$  is a quadratic form in the normal random vector  $\mathbf{x} \sim N[\mu \mathbf{H} \boldsymbol{\theta}, \sigma^2 \mathbf{P}_H]$ . It is known that  $\chi^2/\sigma^2$  is chi-squared distributed with noncentrality parameter  $(\mu^2/\sigma^2) \mathbf{E}_s$ ,  $\mathbf{E}_s = \boldsymbol{\theta}^T \mathbf{H}^T \mathbf{H} \boldsymbol{\theta}$ :  $\chi^2/\sigma^2 \sim \chi_p^2(\mu^2 \mathbf{E}_s/\sigma^2)$ .

The chi-squared distribution has a monotone likelihood ratio. Therefore by the Karlin-Rubin theorem, the test

$$\phi(\chi^2/\sigma^2) = \begin{cases} 1, & \chi^2/\sigma^2 > \chi_0^2 \\ 0, & \chi^2/\sigma^2 \leq \chi_0^2 \end{cases} \quad (5)$$

is the Uniform Most Powerful (UMP) invariant detector for testing  $H_0: \mu=0$  versus  $H_1: \mu>0$  in the measurement  $\mathbf{x} \sim N[\mu \mathbf{H} \boldsymbol{\theta}, \sigma^2 \mathbf{I}]$ . Further we will consider a subspace  $\langle \mathbf{H} \rangle$  as a symmetrical to the group of generalized (logical) shift transformations. Further we establish that statistic (3) is also maximal invariant to the group transformation of general shift for symmetrical signal subspace  $\langle \mathbf{H} \rangle$ .

## 2. DESCRIPTION OF THE SYMMETRICAL SUBSPACE

The operation  $t \oplus \tau$  is called generalized (logical) shift in an

argument  $t$  on a value  $\tau$ , where  $\tau, t \in [0, N-1]$ ,  $t = \sum_{p=1}^n t_p m^{p-1}$ ,

$$\tau = \sum_{p=1}^n \tau_p m^{p-1}, \quad t \oplus \tau = \sum_{p=1}^n (t_p + \tau_p) m^{p-1} = \sum_{p=1}^n c_p m^{p-1},$$

$c_p = ((t_p + \tau_p))_m$  - residue (mod  $m$ ) and  $c_p, t_p, \tau_p \in [0, m-1]$ ,  $N = m^n$ . Let  $\mathbf{g}(\mathbf{h})$  denote the operator of a generalized shift [2]. We represent a discrete mode of a signal as a column vector  $\mathbf{h} = (h_0 \ h_1 \ \dots \ h_{N-1})^T$ .

The generalized shift operation can be represented as permutation of coordinates of this vector. It is possible to represent the operators of generalized shift by block cyclic matrixes of permutations. The matrix  $\mathbf{g}_i \in G$  is a matrix of permutation, therefore one unit is equal to each of its rows and in each column there is only a singular 1, all of the remaining numbers are zero. Let  $\langle \mathbf{H} \rangle$  be a symmetrical subspace. Then  $\mathbf{h}_i = \mathbf{g}_i \mathbf{h}_k$ ,  $i = l \oplus k$ ;  $i, l, k \in [0, N-1]$ . Therefore symmetrical matrix  $\mathbf{H}$  may be written as

$$\mathbf{H} = [\mathbf{h}_0 \ \mathbf{g}_1 \mathbf{h}_0 \ \mathbf{g}_2 \mathbf{h}_0 \ \dots \ \mathbf{g}_{N-1} \mathbf{h}_0]. \quad (6)$$

The subspace  $\langle \mathbf{H} \rangle$  is called symmetrical, if transformed mode  $\mathbf{h}_i \in \mathbf{H}$  by group  $G$  also belongs to subspace  $\langle \mathbf{H} \rangle$ , but mode has another value of the parameter  $i$ :  $\mathbf{g}_i \mathbf{h}_i = \mathbf{h}_i \in \mathbf{H}$ ,  $i, r \in [0, N-1]$ . Note, that  $\mathbf{g}$  is orthogonal matrix:  $\mathbf{g} \mathbf{g}^T = \mathbf{I}$ . We have the following representation for the operator  $\mathbf{g}$ :  $\mathbf{g}_i = \mathbf{V}^H \mathbf{W}_i \mathbf{V}$ ,

where  $\mathbf{V} = N^{-1/2} [\text{Had}(t, \tau)]$ ,  $\text{Had}(t, \tau) = \exp[j2\pi/m \sum_{i=1}^n t_i \tau_i]$ ,

$j = \sqrt{-1}$ ,  $\mathbf{W}_i = \text{diag}[\text{Had}(i, \tau)]$ ,  $\mathbf{V}^H \mathbf{V} = \mathbf{V} \mathbf{V}^H = \mathbf{I}$ . We simplify our notation by written  $(\mathbf{V}^T)^*$  as  $\mathbf{V}^H$ , where  $T$  is sign of transposition,  $*$  - sign of the complex conjugate. Eigenvector of generalized shift operators are the full orthonormalized systems of Hadamard-Chrestenson functions:

$$\text{Had}(p, t) = \exp[j2\pi/m \sum_{i=1}^n p_i t_i], \quad p = \sum_{i=1}^n p_i m^{i-1},$$

$$t = \sum_{i=1}^n t_i m^{i-1}. \quad (7)$$

At  $m=2$  they are called Walsh functions, at  $m=N$  they are called discrete exponential functions.

The matrix  $\mathbf{H}$  is block circulant matrix and may be written as  $\mathbf{H} = \mathbf{V}^H \mathbf{A} \mathbf{V} = \mathbf{g} \mathbf{H} \mathbf{g}$ , for any  $\mathbf{g} \in \mathbf{G}$ , (8)

then  $\mathbf{P}_H = \mathbf{H}(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T = \mathbf{V}^H \mathbf{\Omega} \mathbf{V}$ , (9) where  $\mathbf{A} = \text{diag}(\lambda_0 \lambda_1 \dots \lambda_{N-1})$ ,  $\lambda_i$  - eigenvalue of matrix  $\mathbf{H}$ ,  $\mathbf{\Omega} = \text{diag}(\varepsilon_0 \varepsilon_1 \dots \varepsilon_{N-1})$ ,  $\varepsilon_i$  - eigenvalue of matrix  $\mathbf{P}_H$ . The terms of the diagonal matrix  $\mathbf{A}$  are a Hadamard-Chrestenson Transformation of the first column  $\mathbf{h}$  of matrix  $\mathbf{H}$ . Similarly the terms of the diagonal matrix  $\mathbf{\Omega}$  are a Hadamard-Chrestenson Transformation of the first column of matrix  $\mathbf{P}_H$ .

## 2. NEW DETECTION ALGORITHM FOR SYMMETRICAL SIGNAL SUBSPACE

The sufficient statistic for the parameter  $\mu$  is (3)

$$\chi^2 = \mathbf{x}^T \mathbf{P}_H \mathbf{x}.$$

The operator  $\mathbf{P}_H$  is block circulant matrix and it may be written as

$$\mathbf{P}_H = \mathbf{H}(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T = \mathbf{g} \mathbf{P}_H \mathbf{g}. \quad (10)$$

Now we will establish that statistic (3) is a maximal invariant to the group transformation of general shift  $\mathbf{G} = \{\mathbf{g}: \mathbf{g}(\mathbf{x}) = \mathbf{g} \mathbf{x} = \mathbf{V}^H \mathbf{W} \mathbf{V} \mathbf{x}\}$  under condition (6,8,9). It is clear that:

$$1. (\mathbf{g} \mathbf{x})^T \mathbf{P}_H \mathbf{g} \mathbf{x} = \mathbf{x}^T \mathbf{P}_H \mathbf{x}. \quad (11)$$

$$2. (\mathbf{x}_1)^T \mathbf{P}_H \mathbf{x}_1 = (\mathbf{x}_2)^T \mathbf{P}_H \mathbf{x}_2 \Rightarrow (\mathbf{x}_1)^T \mathbf{V}^H \mathbf{\Omega} \mathbf{V} \mathbf{x}_1 = (\mathbf{x}_2)^T \mathbf{V}^H \mathbf{\Omega} \mathbf{V} \mathbf{x}_2 \Rightarrow (\mathbf{X}_1)^T \mathbf{\Omega} \mathbf{X}_1 = (\mathbf{X}_2)^T \mathbf{\Omega} \mathbf{X}_2 \Rightarrow \|\mathbf{X}_1 \mathbf{\Omega}^{1/4}\|^2 = \|\mathbf{X}_2 \mathbf{\Omega}^{1/4}\|^2 \Rightarrow \mathbf{x}_1 = \mathbf{g} \mathbf{x}_2, \quad (12)$$

where  $\mathbf{X}_1 = \mathbf{V} \mathbf{x}_1$  and  $\mathbf{X}_2 = \mathbf{V} \mathbf{x}_2$  is a Hadamard-Chrestenson transformation of  $\mathbf{x}$ , a sign  $\|\cdot\|$  is Euclidean norm. The maximal invariant may be written as

$$w_p = (1/\sqrt{N}) \sum_{i=0}^{N-1} h_i \text{Had}^*(p, i) \quad (13)$$

The statistic (3) requires  $N^2$  multiplication operations and  $N^2$  addition operations. The new statistic (11) requires  $N$  multiplication operations and  $N^2$  addition operations for  $m=2$ . In this case it is used Walsh Transformation instead of Hadamard-

Chrestenson transformation.

The statistic (3) has not performance penalties if the signal subspace is symmetrical.

But exact symmetry in signal subspace exists seldom for real signal model. Let consider  $N$  the continuous time cosinusoids of the form  $A_i \cos(\omega_i t + \phi_i)$  are summed to produce the signal  $s(t)$ . If this signal is sampled at the sampling instants  $t=nT$ , then the discrete time signal is:

$$s_n = \sum_{i=0}^{N-1} A_i \cos(\omega_i T n + \phi_i).$$

Typically, such samples are taken over an interval  $[0 \leq t < NT]$  to produce the samples vector  $\mathbf{s} = [s_0 s_1 \dots s_{N-1}]^T$ . The vector of samples  $\mathbf{s}$  may be written as  $\mathbf{s} = \text{Re } \mathbf{H} \boldsymbol{\theta}$ ,

where  $\mathbf{H} = [\mathbf{h}_0 \mathbf{h}_1 \dots \mathbf{h}_{N-1}]$ ,  $\boldsymbol{\theta} = [\theta_0 \theta_1 \dots \theta_{N-1}]^T$ ,  $\mathbf{h}_i = [1 \exp(j\omega_i T) \dots \exp(j\omega_i T(N-1))]^T$ ,  $\theta_i = A_i \exp(j\phi_i)$ ,  $\omega_i = \omega_0 + \omega_i$ . We assume that  $\mathbf{s}$  is an  $N$ -vector that is constructed from a linear combination of linearly independent cosines and sines, provided  $T=1$ ,  $\omega_i = (2\pi/N)i$ ,  $i \in [0, N-1]$ . The mode  $\mathbf{h}_i$  is a complex exponential mode and  $\mathbf{H} \mathbf{H}^H = \mathbf{N} \mathbf{I}$ . The algorithm (11) consists of two parts: coherent detector

$$y_k = (1/\sqrt{N}) \sum_{p=0}^{N-1} [w_p (1/\sqrt{N}) \sum_{i=0}^{N-1} x_i \text{Had}^*(p, i) \text{Had}(k, p)] \quad (14)$$

and energy detector  $\chi^2 = \sum_{k=0}^{N-1} (y_k)^2$ . The test (3) may be written as

$$\chi^2 = \mathbf{x}^T \mathbf{P}_H \mathbf{x} = \mathbf{x}^T \mathbf{H}(\mathbf{H}^T \mathbf{H})^{-1} \mathbf{H}^T \mathbf{x} = \frac{\|\mathbf{e}\|^2}{N \|\mathbf{h}\|^2}, \quad (15)$$

$$\text{where } \mathbf{t} = \mathbf{H}^T \mathbf{x}, \quad \mathbf{e} = \|\mathbf{h}\|^2 \quad (16)$$

The known algorithm (15) and obtained algorithm (11) have difference in their coherent detector (14) and (16). We compare signal-to-noise ratio  $(\text{SNR})_1$  for test (14) and  $(\text{SNR})_2$  for test (16) for each mode of  $\mathbf{H}$ . Let  $Z_k = [(\text{SNR})_1]_k / [(\text{SNR})_2]_k$  denote factor of noise immunity loss.  $[(\text{SNR})_1]_k$  may be written as  $[(\text{SNR})_1]_k =$

$$\frac{\mu y_k}{\sigma}, \text{ and } [(\text{SNR})_2]_k \text{ may be written as } [(\text{SNR})_2]_k = \mu N / \sigma. \text{ Then}$$

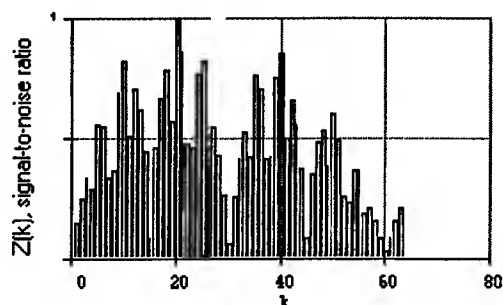
$$\text{factor of noise immunity loss } Z_k = \mu y_k / \mu N. \quad (17)$$

It is plotted for  $N=64$ ,  $M=2$ ,  $\omega_0=1$  in Figure 1. This curve may be used to compute the effective loss in SNR that results from not existing exactly symmetry in a subspace  $\langle \mathbf{H} \rangle$  to the dyadic shift group.

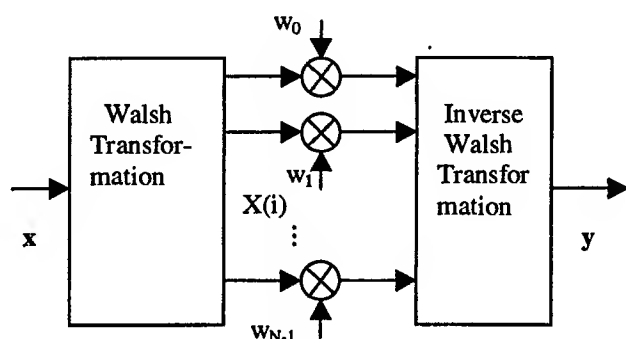
This implementation of coherent detector has  $N$  operations of multiplication and  $N^2$  operations of addition. The structure of implementation for known test  $\mathbf{t} = \mathbf{H}^T \mathbf{x}$  consists of  $N$  branches. Each branch is a correlator of transformed data with stored modes. Therefore known test structure has  $N^2$  operations of multiplication and  $N$  operations of additions. The advantage of new algorithm is obvious. The implementation is hardware-efficient, but it is sub-optimum.

The accuracy of the symmetry in subspace  $\langle \mathbf{H} \rangle$  defines the noise immunity of this algorithm.





**Figure 1** Signal-to-noise effective loss versus mode for  $m=2$  and  $N=64$ ,  $\omega_0 = 1$ .



**Figure 2** Implementation of the coherent detector for symmetrical signal subspace

The accuracy of the symmetry in subspace  $\langle \mathbf{H} \rangle$  defines the noise immunity of this algorithm. In our case the noise immunity losses smaller than 3 dB ( $0.5 + 1$ ) for half of the modes. Our researches have shown that this relation is saved at increase  $N$ . Note that when symmetry in subspace  $\langle \mathbf{H} \rangle$  is not exact, SNR for some modes may be maximized by choosing  $h_0(\omega_0)$ . It is illustrated in Figure 3 for  $m=2$ ,  $N=64$  and  $\omega_0=1.3$ . In this case another some modes have much more SNR than for  $\omega_0=1$  (Fig.1). Note that it is possible to change the type of symmetry in this problem. We can choose  $m=3, 4, 5, \dots$ . But if increasing of  $m$  the complexity of test is increased.

#### 4. SUMMARY

The new algorithm for matched symmetrical subspace detector has been presented. It may be used to reduce the complexity of known algorithm for the signal subspace detection. High quality performance is obtained for moving-target indication under unknown Doppler frequency. The used the new approach for creation of the sub-optimal detector with minimal computing expenditures.

#### 5. REFERENCES

- [1] Louis L. Scharf. *Statistical Signal Processing: Detection, Estimation, and Time Series Analysis*. Addison-Wesley, 1991.
- [2] V.Golikov. "The Theory of Optimal  $M$ -ary Interperiod Processing when Detecting Fluctuating Signals on a Background of Correlated Interference and Noise", *Radioelectronics and Communications System*, vol. 31, pages 2-6, April, 1988.

# MULTIPLE SOURCE DIRECTION FINDING WITH AN ARRAY OF M SENSORS USING TWO RECEIVERS

*E. Fishler and H. Messer*

Department of Electrical Engineering–Systems, Tel Aviv University,  
Tel Aviv 69978, Israel,  
E-mail: : {eranf,messer}@eng.tau.ac.il.

## ABSTRACT

Multiple source direction finding algorithms (e.g., *MUSIC*) are applied on simultaneous measurements collected by  $M$  sensors. However, practical considerations may dictate using less receivers than sensors, such that the measurements cannot be collected simultaneously. In such cases, data is collected sequentially from the different array elements in a process which is referred to as "time varying preprocessing", or "switching".

In this paper we study multiple source direction finding (DF) with an array of  $M > 2$  elements, where only two receivers are available.

## 1. INTRODUCTION AND PROBLEM FORMULATION

Direction finding with fewer receivers than sensors via time varying processing is a very important issue (e.g., [3]). In many practical scenarios the number of receivers is considerably less than the number of sensors. Moreover, the tendency is to use the minimum number of receivers possible which maintain spatial capacity, i.e., only two receivers. Reducing the number of receivers results in a cheaper and simpler design, in the cost of a reduced performance. In this paper we investigate the multiple source localization performance from the identification point of view. We first find how many sources can be localized with only two receivers and then we suggest a computationally efficient algorithm to perform this task.

Assume  $q$  far-field narrow band sources impinging on an array with  $p > q$  sensors from directions  $\{\theta_1, \dots, \theta_q\}$ . Using complex signal representation, the vector of received signals can be written as:

$$\mathbf{x}(t) = \mathbf{A}(\boldsymbol{\theta})\mathbf{s}(t) + \mathbf{n}(t) \quad (1)$$

where  $\mathbf{s}(t)$  is the complex envelope of the slowly varying signals,  $\mathbf{n}(t)$  is the additive noise,  $\boldsymbol{\theta}$  is the vector of directions of arrival, and  $\mathbf{A}(\boldsymbol{\theta}) = [\mathbf{a}(\theta_1), \dots, \mathbf{a}(\theta_q)]$  where  $\mathbf{a}(\theta)$

is the array steering vector at direction  $\theta$ . We denote by  $[\mathbf{x}(t)]_i$  the  $i$ -th element of vector  $\mathbf{x}(t)$ .

Under the standard assumptions about the noise being Gaussian and white and of the signals being Gaussian, the correlation matrix of  $\mathbf{x}(t)$ , denoted by  $R_x(\boldsymbol{\theta})$ , is given by:

$$R_x(\boldsymbol{\theta}) = \mathbf{A}(\boldsymbol{\theta})R_S\mathbf{A}^H(\boldsymbol{\theta}) + \sigma^2\mathbf{I} \quad (2)$$

where  $(\cdot)^H$  denote the complex conjugate transpose operation,  $\sigma^2$  is the noise level and  $R_S$  is the signal covariance matrix.

The problem of estimating  $\boldsymbol{\theta}$  from a set of  $N$  snapshots of the array,  $\mathbf{x}(t_1), \dots, \mathbf{x}(t_N)$ , is usually referred to as the localization problem. The case of spatial samples which are time dependent linear transformation of the array output is discussed in [3]. The resulting model for the measurements is  $\mathbf{y}(t_i) = \mathbf{G}(t_i)\mathbf{x}(t_i)$ , where  $\mathbf{G}(t_i)$  is the time dependent linear transformation. Note that  $\mathbf{G}(t_i)$  is a matrix in which the number of rows is the number of receivers used at time  $t_i$ .

We are interested in the special case where  $\mathbf{G}(t_i)$  is a  $2 \times p$  matrix such that each row is a vector with all elements but one equal zero, where the non zero element equals 1. Without loss of generality, we assume that we take  $N$  snapshots of each sub array of two elements. The total number of snapshots is  $L = \binom{p}{2}N$ . At time instant  $t_i$ ,  $i = 1, \dots, L$ , the output of the reduced array is:  $\mathbf{y}(t_i) = [[\mathbf{x}(t_i)]_k [\mathbf{x}(t_i)]_l]^T$  for some  $k \neq l \in \{1, \dots, p\}$ .

## 2. SOME RELATED RESULTS

In [3] the  $ML$  estimator for a general transformation matrix  $\mathbf{G}(t_i)$  is presented. This procedure involves maximization over all unknown parameters:  $\boldsymbol{\theta}, \sigma^2, R_S$ . This maximization problem becomes extremely difficult even for as little as two sources. The authors presented an ad-hoc approach, the *GLS*, which reduces the complexity of the estimator to a search over only  $q$  parameters.

Alternatively, by noting that our problem can be modeled as a problem of direction finding with time-varying array, one can apply the results of [4] which include, among

others, expressions for the Cramer Rao lower bound (CRLB) on the estimation error of the unknown parameters. Also, some conjectures about the complexity of the ML estimator were presented which suggested that in the general case the ML estimator is not separable.

In [2] it is shown that, unlike the case where the array is sampled simultaneously, in cases where the number of sensors in the sub-array is smaller than the number of sources, the CRLB for  $\theta_1, \dots, \theta_q$  does not approach zero as the SNR approaches infinity, so the time varying spatial sampling process causes a residual estimation error.

Eigenvector based methods for the case of time-varying arrays had been proposed in [1]. In this paper two possible eigenvector based method have been proposed. One is based on an interpolating matrix and the other is based on a focus matrix. However, both methods can not be applied to our problem due to the large differences in the steering vectors between successive time instances.

### 3. THE IDENTIFIABILITY PROBLEM

It is well known that when the array is simultaneously sampled so (1) holds, and under some very weak conditions on the array, one can localize up to  $p - 1$  sources. Is it also true when only two receivers are used? The following theorem refers to this question:

**Theorem 1** *Using an array of  $p$  sensors and only two receivers, up to  $q = p - 1$  narrowband sources can be uniquely localized.*

**Proof 1** *Let  $\tilde{\mathbf{y}}(t_i)$  be a column vector with  $2\binom{p}{2}$  elements, given by:*

$$\tilde{\mathbf{y}}(t_i) = [\mathbf{y}(t_i)^T, \mathbf{y}(t_{i+N})^T, \mathbf{y}(t_{i+2N})^T, \dots, \mathbf{y}(t_{i+(L-N)})^T]^T \quad (3)$$

*Without loss of generality, assume that first we take  $N$  samples of the first and second sensors simultaneously. Next we take another  $N$  samples from the first and third sensors simultaneously, and so on.  $\tilde{\mathbf{y}}(t_1)$  is a column vector, with the first two elements equal to the first sample of the first two sensors sampled. The third or fourth elements of  $\tilde{\mathbf{y}}(t_1)$  are the two elements of the first sample from the second and third sensors, and so on. It is clear that  $\{\tilde{\mathbf{y}}(t_i)\}_{i=1}^N$  contain all the available samples and thus it contains all the statistical information on the unknown parameters.*

*It can be easily verified that  $\{\tilde{\mathbf{y}}(t_i)\}_{i=1}^N$  are i.i.d. complex Gaussian vectors with block diagonal correlation matrix,  $R_{\tilde{\mathbf{y}}}(\underline{\theta})$ , given by*

$$[R_{\tilde{\mathbf{y}}}(\underline{\theta})]_{ij} = \begin{cases} 0 & |i - j| > 1 \\ [R_x(\underline{\theta})]_{kl} & i > j \\ [R_x(\underline{\theta})]_{lk} & i < j \\ [R_x(\underline{\theta})]_{kk} & i = j, \lfloor \frac{i}{2} \rfloor \neq \frac{i}{2} \\ [R_x(\underline{\theta})]_{ll} & O.W \end{cases} \quad (4)$$

where  $k$  and  $l$  are the first and second sensors sampled at the  $\lfloor \frac{i}{2} \rfloor$  switching. It is clear from the structure of  $R_{\tilde{\mathbf{y}}}$  that a simple one to one mapping, denoted by  $\psi(R_x)$ , between  $R_x$  and  $R_{\tilde{\mathbf{y}}}$ , exists

Let  $\underline{\theta} = [\theta_1, \dots, \theta_k]$  and  $\underline{\theta}' = [\theta_1, \dots, \theta_{k'}]$  be two sets of bearings, such that  $k, k' \leq q - 1$  and  $\underline{\theta}' \neq \underline{\theta}$ . For the case of simultaneous sampling up to  $q - 1$  sources could be uniquely localized, i.e.,  $R_x(\underline{\theta}) \neq R_x(\underline{\theta}')$  for every  $\underline{\theta} \neq \underline{\theta}'$ . Now, using the fact that  $\psi$  is a one to one mapping between  $R_x$  and  $R_{\tilde{\mathbf{y}}}$ , it is clear that  $R_{\tilde{\mathbf{y}}}(\underline{\theta}) \neq R_{\tilde{\mathbf{y}}}(\underline{\theta}')$  for every  $\underline{\theta} \neq \underline{\theta}'$ .

In addition, since  $\tilde{\mathbf{y}}(t_i)$  is a complex Gaussian vector, the p.d.f. of  $\tilde{\mathbf{y}}(t_i)$  given  $\underline{\theta}$  is different from the p.d.f. of  $\tilde{\mathbf{y}}(t_i)$  given  $\underline{\theta}'$ , which is a sufficient condition for identifiability.

This theorem provides a very important result: at each time instant we are sampling a sub array of size two which in turn enable us to localize only one source. However, coherently combining all the results from the sub arrays, enables one to localize  $p - 1$  sources, the same number as if we were sampling the all array with  $p$  receivers.

### 4. EIGENVECTORS BASED METHODS

The ML estimator for  $\theta$  requires a  $q$  dimensional search, at least. Eigenvector based methods, like the MUSIC, offers a way to reduce the complexity to a one dimensional search. This reduction in complexity is crucial since, still today, with the most advanced DSP, searching in more than two dimensional space can not be performed in real time.

We next describe a new eigenvector based procedure which can be used in our problem. We start with the following equivalent description of the data:

Let  $\mathbf{z}(t_i)$  be a column vector with  $p$  elements. Let all the elements be equal zero except, say the  $k$  and  $l$  elements, which are equal to  $[\mathbf{x}(t_i)]_k$  and  $[\mathbf{x}(t_i)]_l$ , respectively. That is,  $k, l$  are the two array elements which are sampled at time  $t_i$ . Now, denote by  $\hat{R}_Z = \frac{1}{N} \sum_{i=1}^L \mathbf{z}(t_i) \mathbf{z}^H(t_i)$  the empirical correlation matrix, it can be shown that its expected value is given by:

$$R_Z = \mathbf{A}(\theta) R_S \mathbf{A}^H(\theta) + \sigma^2 \mathbf{I} + \mathbf{\Lambda} \quad (5)$$

where  $\mathbf{\Lambda}$  is a diagonal matrix whose diagonal entries are  $(p - 1) \cdot \text{diag}(\mathbf{A}(\theta) R_S \mathbf{A}^H(\theta) + \sigma^2 \mathbf{I})$ . The matrix  $\mathbf{\Lambda}$  is the only difference from the mean of the sample covariance matrix in the case where the array is sampled simultaneously, where eigenvalue based methods are easily applied, and the mean of the sample covariance matrix where only two receivers are used simultaneously.

However, if all the elements of the diagonal matrix  $\mathbf{\Lambda}$  are equal, then eigenvector based methods for estimating  $\theta$  can still be used, since it is just added to the noise covariance

matrix so it effectively changes the (unknown) noise level. There are two sufficient conditions for all the elements of  $\mathbf{A}$  to be equal:

1. All sources are uncorrelated, so  $R_S$  is a diagonal matrix.
2. All the array elements are omnidirectional, such that  $|\mathbf{a}_i(\theta)| = |\mathbf{a}_j(\theta)| \forall i \neq j$  and for any  $\theta$ .

However, since these conditions are rarely fully fulfilled in practice, MUSIC like procedures cannot be applied on  $R_Z$  directly.

A careful examination of  $R_Z(\theta)$  and of  $R_X(\theta)$  shows that their off-diagonal elements are the same, while the diagonal elements of  $R_Z(\theta)$  are  $p-1$  larger,  $\forall \theta$ . We therefore suggest a non-linear pre-processing procedure: to divide the diagonal elements of  $\hat{R}_Z$  by  $p-1$ . Denote by  $\hat{R}_Z$  the resulting matrix, it can be easily verified that  $E\{\hat{R}_Z\} = R_X(\theta)$  and thus  $\hat{R}_Z$  can be used with all the eigenvector based methods, e.g. MUSIC. We refer to the MUSIC with the suggested preprocessing as *MMUSIC*. Naturally, the performance of the *MUSIC* and of the *MMUSIC* applied to the same array will be different, since only the first moment (the expected value) of  $\hat{R}_X$  and of  $\hat{R}_Z$  is the same.

This method can be extended to cases where the number of samples taken from each sensor is not equal. Let  $n_i$  be number of samples taken at the  $i$ -th switching. Let  $\hat{R}_Z = \sum \mathbf{z}(t_i)\mathbf{z}^H(t_i)$ . It can be verified that the mean of  $\hat{R}_Z$  is given by:

$$E\{\hat{R}_Z\} = (\mathbf{A}(\theta)R_S\mathbf{A}(\theta)^H + \sigma^2\mathbf{I}) \odot \Psi \quad (6)$$

where  $(\Psi)_{ij}$  is the total number of snapshots taken from the  $i, j$  sensors simultaneously,  $(\Psi)_{i,i}$  is the total number of snapshots taken from  $i$ -th sensor, and  $\odot$  denotes element by element matrix multiplication. The suggested preprocessing in this case is to divide each element of  $\hat{R}_Z$  by the corresponding element in  $\Psi$ . The resulting matrix, denoted again by  $\hat{R}_Z$ , can be used with any eigenvalue based method.

## 5. SIMULATION STUDY

Consider a uniform linear array with 4 omni-directional elements. Assume two equi-power, partially correlated ( $\rho = 0.25$ ) sources at bearings  $0^\circ, 15^\circ$  and  $N = 100$ . In Figure 1 a typical spectrum of the *MMUSIC* is shown. For comparison, we show a typical spectrum of the *MUSIC* which has been applied on  $\hat{R}_Z$  without preprocessing. It shows that without preprocessing the two sources are not resolved, so, as predicted, the *MUSIC* cannot be used directly for multiple source localization.

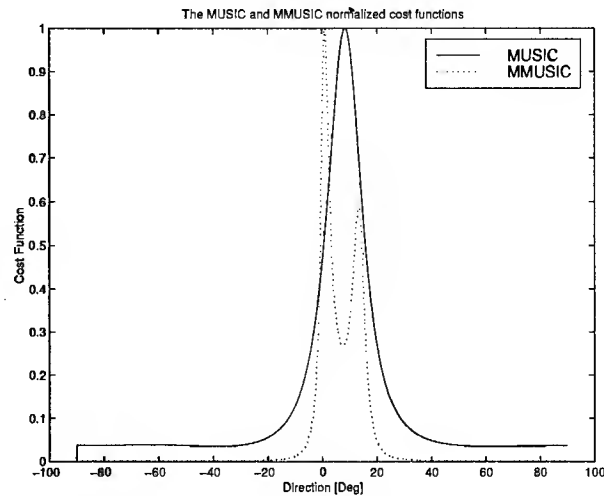


Figure 1: Typical *MUSIC* and *MMUSIC* cost functions

We now present results of a simulation performance study for the same experiment. Figures 2 and 3 depict the probability of detecting two sources and the *MSE* of the bearing of the first source, respectively, for various correlation coefficients, as a function of the SNR. These results are based on averaging of 1000 Monte Carlo Runs.

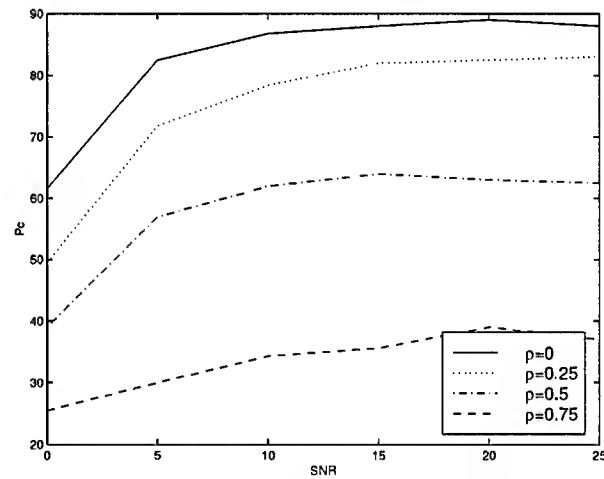


Figure 2: The probability of detecting two sources as a function of the SNR.

Figures 4 and 5 depict the probability of detecting two sources and the *MSE* of the bearing of the first source, respectively, as a function of the number of snapshots, where the SNR is fixed at 10 dB.

Generally speaking, this study suggests that the performance of the *MMUSIC* improves as the SNR increases, as the number of snapshots increases and as the correlation between the sources decreases. However, our future work will focus on analytic performance analysis of the algorithm

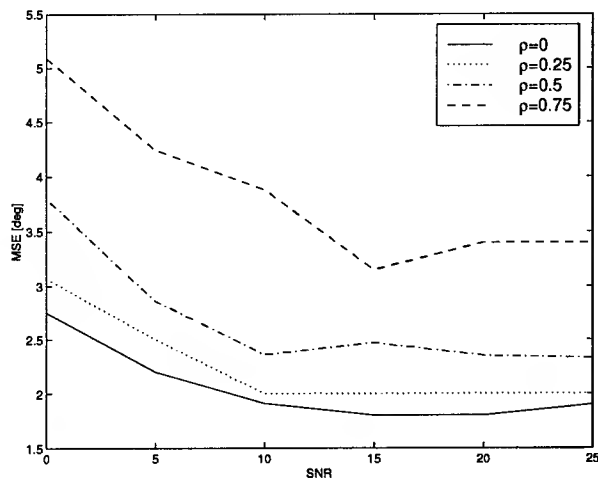


Figure 3: The MSE of the bearing of the first source as a function of the SNR.

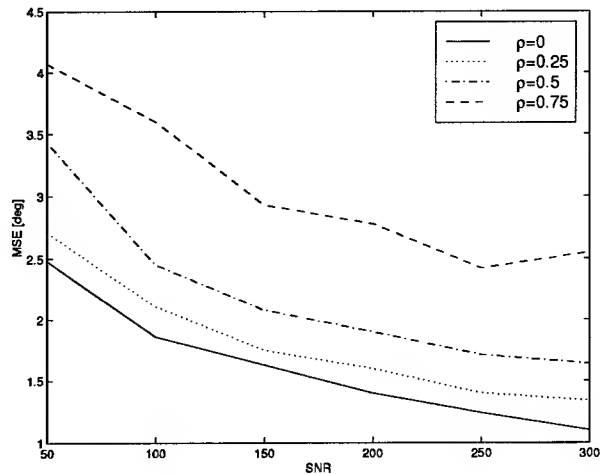


Figure 5: The MSE of the bearing of the first source as a function of the number of snapshots.

so its inherent limitations can be exploited.

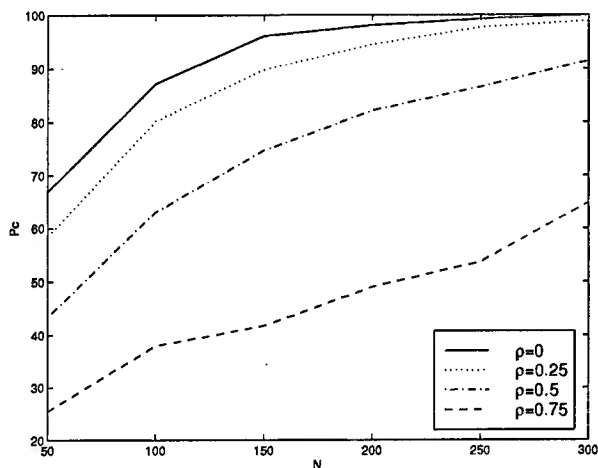


Figure 4: The probability of detecting two sources as a function of the number of snapshots.

## 6. REFERENCES

- [1] B. Friedlander and A. Zeira, "Eigenstructure-based algorithms for direction finding with time-varying arrays," *IEEE Trans. on AES*, Vol. 32, pp. 689 - 701, April 1996.
- [2] J. Sheinvald, "On Detection and Localization of Multiple Signals by Sensor Arrays," Ph.D. Dissertation, Tel Aviv University, Israel.
- [3] J. Sheinvald and M. Wax, "Direction Finding with

Fewer Receivers via Time-Varying Preprocessing", *IEEE Trans. on SP*, Vol. 47, pp. 2-10, January 1999.

- [4] A. Zeira and B. Friedlander, "Direction Finding with Time Varying Arrays", *IEEE Trans. on SP*, Vol. 43, pp. 927 - 938, 1995.
- [5] M. A. Doron, A. J. Weiss and H. Messer, "Maximum Likelihood direction finding of wide band sources", *IEEE Trans. on SP*, Vol. 41, pp. 411 - 414, 1993.

# SELF-STABILIZED MINOR SUBSPACE EXTRACTION ALGORITHM BASED ON HOUSEHOLDER TRANSFORMATION

K. Abed-Meraim, S. Attallah\*\*, A. Chkeif\*, Y. Hua\*\*\*

\* Telecom Paris, TSI Dept. 46, rue Barrault, 75634, Paris Cedex 13 France.

\*\* Centre for wireless communication, National University of Singapore, Singapore.

\*\*\* The University of Melbourne, Elec. Eng. Dept. Parkville, Vic. 3052, Australia.

E-mails: abed,chkeif@tsi.enst.fr, cwcsa@leonis.nus.edu.sg, yhua@ee.mu.oz.au.

## ABSTRACT

In this paper, we propose an orthogonalized version of OJA algorithm (OOJA) that can be used for the estimation of minor and principal subspaces of a vector sequence. The new algorithm offers, as compared to OJA, such advantages as orthogonality of the weight matrix, which is ensured at each iteration, numerical stability and a quite similar computational complexity.

## 1. INTRODUCTION

Principal and minor component analysis (PCA and MCA), which are part of the more general principal and minor subspace (PSA and MSA) analysis, are two important problems that are frequently encountered in many information processing fields.

Let  $\{r(k)\}$  be a sequence of  $N \times 1$  random vectors with covariance matrix  $C = E[r(k)r^T(k)]$ . Consider the problem of extracting the principal or the minor subspace spanned by the sequence, of dimension  $P < N$ , assumed to be the span of the  $P$  principal or minor eigenvectors of the covariance matrix, respectively. To solve this problem, several subspace extraction algorithms have so far been proposed [1]-[5]. The minor subspace extraction algorithm of Oja *et al.* [4] can be formulated as

$$\begin{aligned} W(i+1) &= W(i) - \beta [r(i)y^T(i) - W(i)y(i)y^T(i))] \\ &= W(i) - \beta p(i)y^T(i) \end{aligned} \quad (1)$$

where  $W(i) \in \mathbb{R}^{N \times P}$  is the minor subspace estimate,  $y(i) \triangleq W^T(i)r(i)$ ,  $p(i) \triangleq (r(i) - W(i)y(i))$ , and  $\beta > 0$  is a learning parameter. Reversing the sign of the adaptive gain, i.e., replacing  $-\beta$  in (1) by  $+\beta$ , yields a principal subspace extraction algorithm. Chen *et al.* have proposed a novel MSA algorithm [5] which can be written as follows

$$W(i+1) = W(i) - \beta [r(i)y(i)^T W^T(i)W(i)$$

$$- W(i)y(i)y^T(i)] \quad (2)$$

The discrete-time update of (2) suffers from a marginal instability similar to the PCA ( $P = 1$ ) algorithm in [2]. Recently, a novel self-stabilizing MSA algorithm given by

$$\begin{aligned} W(i+1) &= W(i) - \beta [r(i)y(i)^T W^T(i)W(i) \times \\ &\quad W^T(i)W(i) - W(i)y(i)y^T(i)] \end{aligned} \quad (3)$$

has been proposed by Douglas *et al.* in [3].

## 2. ORTHOGONAL OJA

Our algorithm consists of (1) plus an orthogonalization step of the weight matrix to be performed at each iteration. Orthogonality is an important property that is desired in many subspace based estimation methods [6]. To this end, we set (using informal notation):

$$W(i+1) := W(i+1)(W^T(i+1)W(i+1))^{-1/2} \quad (4)$$

where  $(W^T(i+1)W(i+1))^{-1/2}$  denotes an inverse square root of  $(W^T(i+1)W(i+1))$ . To compute the latter, we use the updating equation of  $W(i+1)$ . Keeping in mind that  $W(i)$  is now an orthogonal matrix, we have

$$W^T(i+1)W(i+1) = I + \beta^2 \|p(i)\|^2 y(i)y^T(i) = I + \mathbf{x}\mathbf{x}^T,$$

where we have used the fact that  $W^T(i)p(i) = 0$ ,  $I$  is the identity matrix, and  $\mathbf{x} \triangleq \beta \|p(i)\| y(i)$ . Using

$$(I + \mathbf{x}\mathbf{x}^T)^{-1/2} = I + \left( \frac{1}{\sqrt{1 + \|\mathbf{x}\|^2}} - 1 \right) \frac{\mathbf{x}\mathbf{x}^T}{\|\mathbf{x}\|^2},$$

we obtain

$$(W^T(i+1)W(i+1))^{-1/2} = I + \tau(i)y(i)y^T(i), \quad (5)$$

where  $\tau(i) \triangleq \frac{1}{\|y(i)\|^2} \left( \frac{1}{\sqrt{1 + \beta^2 \|p(i)\|^2 \|y(i)\|^2}} - 1 \right)$ . Substituting (5) into (4) and using the updating equation of  $W(i+1)$  leads to

$$\begin{aligned} W(i+1) &= (W(i) - \beta p(i) y^T(i))(I + \tau(i) y(i) y^T(i)) \\ &= W(i) - \beta \bar{p}(i) y^T(i), \end{aligned} \quad (6)$$

where  $\bar{p}(i) \triangleq -\tau(i) W(i) y(i) / \beta + (1 + \tau(i) \|y(i)\|^2) p(i)$ . Thus, the algorithm can be written as

- Initialization of the algorithm:

$W(0) = \text{any arbitrary orthogonal matrix.}$

- Algorithm at iteration  $i$ :

$$\begin{aligned} y(i) &= W^T(i) r(i) \\ z(i) &= W(i) y(i) \\ p(i) &= r(i) - z(i) \\ \phi(i) &= \frac{1}{\sqrt{1 + \beta^2 \|p(i)\|^2 \|y(i)\|^2}} \\ \tau(i) &= \frac{\phi(i) - 1}{\|y(i)\|^2} \\ \bar{p}(i) &= -\tau(i) z(i) / \beta + \phi(i) p(i) \\ W(i+1) &= W(i) - \beta \bar{p}(i) y^T(i) \end{aligned}$$

In order to gain more insight into OOJA algorithm we must examine the following points:

#### 1. Minor subspace:

In terms of orthogonality errors, OOJA algorithm guarantees the orthogonality of the weight matrix at each iteration. With the orthogonalization the three algorithms (1), (2), and (3) become identical. However, simulation results show that the discrete-time update of OOJA algorithm is sensitive to the propagation of rounding-off errors. Fortunately, we can overcome this problem by reformulating the algorithm equations as shown in section 3.

#### 2. Principal subspace:

With respect to subspace errors, our algorithm converges at the same rate as (1). In terms of orthogonality errors, it guarantees the orthogonality of the weight matrix at each iteration, whereas (1) converges to an orthogonal weight matrix only asymptotically. Finally, it is worth noting that (3) quickly diverges for PSA.

#### 3. Computational complexity:

The computational complexity of algorithms (3)

and (2) are  $7NP + O(N)$  and  $5NP + O(N)$  flops per iteration, respectively. OOJA and (1) cost, however, only  $3NP + O(N)$  flops per iteration. It is interesting to note that the orthogonalization step does not increase the computational cost of OJA algorithm. On the other hand, the updating equation of the weight matrix of OOJA algorithm has a more compact form than (2) and (3), i.e., it uses only one outer product instead of two for (2) and (3). This turns out to be useful when a subspace extraction algorithm is cascaded with other adaptive algorithms, e.g., [8].

#### 4. Convergence:

The convergence of OOJA algorithm follows directly from that of OJA algorithm [7]. In fact, (6) can be rewritten as  $W(i+1) = W(i) - \beta p(i) y^T(i) + O(\beta^2)$ . Therefore, for  $\beta \ll 1$ , it can be shown that the two algorithms have the same convergence performance.

On the other hand the convergence proof of (3) is not complete. Effectively, to prove that  $\text{span}[W]$  converges to  $\text{span}[E_2]$ , where  $E_2$  is the minor  $P$ -dimensional subspace spanned by the eigenvectors corresponding to the  $P$  smallest eigenvalues, Douglas *et al.* [3] have used the following assumption:

If all the eigenvalues of  $M(t)$  have negative real parts, then for the following system

$$\frac{dQ(t)}{dt} = M(t)Q(t),$$

we have

$$\lim_{t \rightarrow \infty} Q(t) = 0.$$

This assumption is true if  $M(t)$  is time invariant but not always true when  $M(t)$  is time variant as shown by the counter examples given in [10, 11].

### 3. IMPLEMENTATION USING HOUSEHOLDER TRANSFORMATION

Because of the numerical instability of OOJA when used for minor subspace estimation, we propose here another implementation of the algorithm based on Householder transformation. In fact, the new implementation can be derived from a reformulation of (6) in terms of Householder transformation. We have the following result:

**Proposition 1** Let  $u(i) \triangleq \bar{p}(i) / \|\bar{p}(i)\|$ . Then equation (6) can be rewritten as

$$W(i+1) = H(i)W(i) \quad (7)$$

where  $\mathbf{H}(i)$  is the Householder transformation given by

$$\mathbf{H}(i) \triangleq \mathbf{I} - 2\mathbf{u}(i)\mathbf{u}^T(i)$$

Based on this result (see appendix for proof), the new implementation consists in computing successively  $\mathbf{y}(i)$ ,  $\mathbf{p}(i)$ ,  $\tau(i)$ , and  $\bar{\mathbf{p}}(i)$ . Then, we compute

$$\begin{aligned}\mathbf{u}(i) &= \bar{\mathbf{p}}(i)/\|\bar{\mathbf{p}}(i)\| \\ \mathbf{v}(i) &= \mathbf{W}^T(i)\mathbf{u}(i) \\ \mathbf{W}(i+1) &= \mathbf{W}(i) - 2\mathbf{u}(i)\mathbf{v}^T(i)\end{aligned}$$

Since the decomposition of the weight matrix involves the use of numerically well-behaved Householder orthogonal matrices (see [9] pp.209-213), OOJA becomes numerically very stable. The new implementation presents now a computational complexity of  $4NP + O(N)$  flops per iteration.

#### 4. SIMULATION RESULTS

**Example 1:** In this example, we choose  $\mathbf{r}(i)$  to be a sequence of independent jointly-Gaussian random vectors with covariance matrix

$$\mathbf{C} = \begin{pmatrix} 0.9 & 0.4 & 0.7 & 0.3 \\ 0.4 & 0.3 & 0.5 & 0.4 \\ 0.7 & 0.5 & 1.0 & 0.6 \\ 0.3 & 0.4 & 0.6 & 0.9 \end{pmatrix} \quad (8)$$

$P = 2$ ,  $\beta = 0.01$ , and as recommended in [5]  $\mathbf{W}(0) = \mathbf{D}$ , where  $\mathbf{D}_{i,j} = \delta(j-i)$ . As in [3], we calculate the ensemble averages of the performance factors

$$\rho(i) = \frac{1}{r_0} \sum_{r=1}^{r_0} \frac{\text{tr}(\mathbf{W}_r^T(i)\mathbf{E}_1 * \mathbf{E}_1^T \mathbf{W}_r(i))}{\text{tr}(\mathbf{W}_r^T(i)\mathbf{E}_2 * \mathbf{E}_2^T \mathbf{W}_r(i))}, \quad (9)$$

$$\eta(i) = \frac{1}{r_0} \sum_{r=1}^{r_0} \|\mathbf{W}_r^T(i)\mathbf{W}_r(i) - \mathbf{I}\|_F^2, \quad (10)$$

where the number of algorithm runs is  $r_0 = 100$ ,  $r$  indicates that the associated variable depends on the particular run,  $\|\cdot\|_F$  denotes the Frobenius norm, and  $\mathbf{E}_1$  (respectively  $\mathbf{E}_2$ ) is the principal  $(N-P)$ -dimensional subspace (respectively minor  $P$ -dimensional subspace). Figure 1 compares the performance of OOJA (without Householder implementation) with (1), (2), and (3). As we can see our algorithm behaves better than (1) and (2), but still suffers from numerical instability.

**Example 2:** In this example all parameters are kept the same as in the first example. Figure 2 shows the performance of Householder-based OOJA algorithm as compared to (1), (2), and (3). We can see that the

new implementation is numerically stable.

**Example 3:** We consider here the same context as in the previous examples. By reversing the sign of  $\beta$ , we extract now the principal  $P$ -dimensional subspace. In (9), we replace  $\mathbf{E}_1$  by  $\mathbf{E}_2$  and vice versa. As we can see from figure 3, our algorithm (without Householder implementation) is numerically stable and has better performance than (1), (2), and (3).

#### 5. CONCLUSIONS

In this paper, we proposed an orthogonal OJA (OOJA) algorithm that can perform both PCA and MCA by simply switching the sign of the same learning rule. We gave two fast implementations of OOJA where the orthogonality of the weight matrix is ensured at each iteration. OOJA is numerically stable and its computational complexity is smaller than those reported in [3] and [5].

#### 6. APPENDIX

*Proof of proposition 1:* Using the definition<sup>1</sup> of  $\mathbf{y}$  we can write  $\bar{\mathbf{p}}\mathbf{y}^T = \bar{\mathbf{p}}\mathbf{r}^T\mathbf{W}$ . By decomposing the observation vector as:

$$\begin{aligned}\mathbf{r} &= \mathbf{W}\mathbf{W}^T\mathbf{r} + (\mathbf{I} - \mathbf{W}\mathbf{W}^T)\mathbf{r} \\ &= \mathbf{W}\mathbf{y} + \mathbf{p} \\ &= \frac{-\beta}{\tau} \left[ \frac{-\tau}{\beta} \mathbf{W}\mathbf{y} + \frac{-\tau}{\beta} \mathbf{p} \right],\end{aligned}$$

we can write

$$\begin{aligned}\bar{\mathbf{p}}\mathbf{r}^T\mathbf{W} &= \frac{-\beta}{\tau} \bar{\mathbf{p}} \left[ \frac{-\tau}{\beta} \mathbf{W}\mathbf{y} + \frac{-\tau}{\beta} \mathbf{p} \right]^T \mathbf{W} \\ &= \frac{-\beta}{\tau} \bar{\mathbf{p}} \left[ \frac{-\tau}{\beta} \mathbf{W}\mathbf{y} + (1 + \tau\|\mathbf{y}\|^2)\mathbf{p} \right]^T \mathbf{W} \\ &= \frac{-\beta}{\tau} \bar{\mathbf{p}}\bar{\mathbf{p}}^T\mathbf{W}.\end{aligned}$$

where the second equality comes from the fact that  $\mathbf{p}^T\mathbf{W} = \mathbf{0}$ . Finally, we obtain  $\mathbf{W}(i+1) = (\mathbf{I} + \frac{\beta^2}{\tau(i)}\bar{\mathbf{p}}(i)\bar{\mathbf{p}}(i)^T)\mathbf{W}$ . To complete the proof we have to show that

$$\frac{\beta^2}{\tau} = \frac{-2}{\|\bar{\mathbf{p}}\|^2} \quad \text{or equivalently} \quad \|\bar{\mathbf{p}}\|^2 = \frac{-2\tau}{\beta^2}.$$

Using the definition of  $\bar{\mathbf{p}}$  and the equality  $1 + \tau\|\mathbf{y}\|^2 = (1 + \beta^2\|\mathbf{p}\|^2\|\mathbf{y}\|^2)^{-1/2}$ , we can write

$$\|\bar{\mathbf{p}}\|^2 = \frac{\tau^2\|\mathbf{y}\|^2}{\beta^2} + \frac{\|\mathbf{p}\|^2}{1 + \beta^2\|\mathbf{p}\|^2\|\mathbf{y}\|^2}$$

<sup>1</sup>Here, we omit the time index  $i$  to simplify the notations.



$$\begin{aligned}
&= \frac{\tau^2 \|y\|^2}{\beta^2} + \frac{1}{\beta^2 \|y\|^2} \left(1 - \frac{1}{1 + \beta^2 \|p\|^2 \|y\|^2}\right) \\
&= \frac{1}{\beta^2 \|y\|^2} ((\tau \|y\|^2)^2 + 1 - (1 + \tau \|y\|^2)^2) \\
&= \frac{-2\tau}{\beta^2} \quad \square
\end{aligned}$$

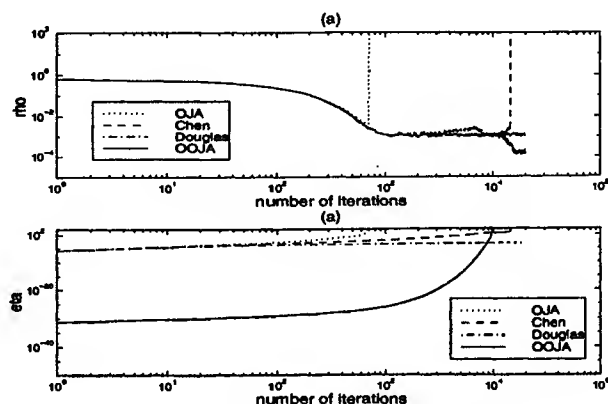


Figure 1: Average behaviors for MSA.

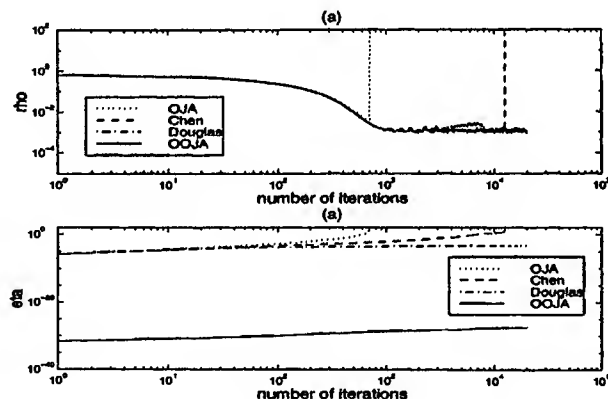


Figure 2: Average behaviors for MSA using Householder-based implementation.

## 7. REFERENCES

[1] Y. Hua, Y. Xiang, T. Chen, K. Abed-Meraim, and Y. Miao, "A New Look at the Power Method for Fast Subspace Tracking", *Digital Signal Processing, Academic Press*, Oct. 1999.

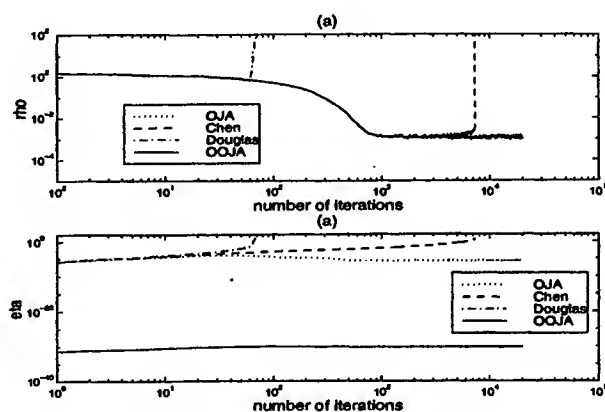


Figure 3: Average behaviors for PSA.

- [2] T. P. Krasulina, "Method of Stochastic Approximation in the Determination of the Largest Eigenvalue of the Mathematical Expectation of Random Matrices," *Automat. Remote Contr.*, vol 2, pp. 215-221, 1970.
- [3] S. C. Douglas, S.-Y. Kung, and S.-I. Amari, "A Self-Stabilized Minor Subspace Rule," *Sig. Process. Letters*, vol. 5, no. 12, pp. 328-330, Dec. 1998.
- [4] E. Oja, "Principal Components, Minor Components, and Linear Neural Networks," *Neural Networks*, vol. 5, pp. 927-935, Nov./Dec. 1992.
- [5] T. Chen, S.-I. Amari, and Q. Lin, "A Unified Algorithm for Principal and Minor Components Extraction," *Neural Networks*, vol. 11, pp. 385-390, 1998.
- [6] S. Marcos, A. Marsal, and M. Benidir, "The Propagator Method for Source Bearing Estimation," *Sig. Proc.*, vol 42, pp. 121-138, Apr. 1995.
- [7] T. Chen, Y. Hua, and W. Yan, "Global Convergence of Oja's Subspace Algorithm for Principal Component Extraction," *IEEE Trans. on Neural Network*, pp. 58-67, Jan. 1998.
- [8] A. Chkeif, K. Abed-Meraim, G. Kawas Kaleh, and Y. Hua, "Blind Adaptive Multiuser Detection With Antenna Array," *accepted for publication in IEEE Trans. on Comm.*
- [9] G. H. Golub and C. F. Van loan, "Matrix computations," the Johns Hopkins University Press, 1996.
- [10] L. Markus and H. Yamabe, "Global Stability Criteria for Differential Systems" *J. Osaka Math.*, vol. 12, pp. 305-317, 1960.
- [11] R. E. Vinograd, "Remark on the Critical Case of Stability of a Singular Point in the Plane" *Doklady Akad. Nauk*, vol. 101, pp. 209-212, 1955.

# A BOOTSTRAP TECHNIQUE FOR RANK ESTIMATION

*Per Pelin, Ramon Brcich and Abdelhak Zoubir*

Australian Telecommunications Research Institute<sup>1</sup>(ATRI), Curtin University of Technology,  
GPO Box U 1987, Perth WA 6845, Australia. E-mail: pelle@atri.curtin.edu.au

## ABSTRACT

A crucial step in many signal processing applications is the determination of the effective rank of a noise corrupted multi-dimensional signal, i.e., the dimension of the signal subspace. Standard techniques for rank estimation, such as the minimum description length, often have shortcomings in practice, an example being when noise parameters are unknown. An alternative scheme is proposed for rank detection. From successive pairs of the ordered eigenvalues of the array covariance, a series of statistics is formed. The statistics are chosen such that their distributions for noise eigenvalue pairs are close. The actual distributions are unknown and are estimated with the Bootstrap. The rank is then found by a sequential comparison of the estimated distributions using a Kolmogorov-Smirnov test.

## 1. INTRODUCTION

Many signal processing algorithms, such as direction finding algorithms, rely on the low-rank structure of a multi-dimensional signal. The rank typically has an interpretation as the model order, revealing the number of signals hidden in noise, or the dimension of a low-order signal subspace. Therefore, finding the effective rank of a noise corrupted signal is a crucial initial step in many applications.

Classical techniques to estimate the rank when the noise is Gaussian include the minimum description length (MDL) and Akaike's information theoretic criterion (AIC) [10], and their subjective counterpart the sphericity test [2]. In the latter, a threshold is set to obtain a desired level of the test, whereas in the objective MDL and AIC, the actual threshold is dependent on the data size by asymptotic arguments. Nevertheless, they all rely on the structure of the noise eigenvalues of the covariance matrix, and it is required that the actual spatial noise color is known. If the noise assumptions are violated, for example, when the noise has an unknown spatial color, detection performance is degraded. For noise of unknown color, an alternative to eigenvalue-based tests is to use properties of canonical correlations [2], as in [11][12]. However, these schemes put some restrictions on the structure of the data model, limiting their applicability.

1. This work was in part supported by the Australian Telecommunications Cooperative Research Centre (AT-CRC).

To mitigate the problem of slight uncertainties in the noise model, both w.r.t. possible non-Gaussianity and noise color, a new technique for rank detection is proposed. The detection procedure is based on a property of the marginal distributions of the noise sample eigenvalues. Instead of relying on parametric assumptions, these distributions are estimated from the data using the Bootstrap [5]. Based on these estimates, the distributions of a series of secondary variables are estimated, on which the actual rank estimation is performed using a robust Kolmogorov-Smirnov test [7]. The necessary number of Bootstrap resamples is surprisingly small, keeping the computational cost at a reasonable level.

## 2. MODELING

Consider  $m$ -variate data according to the linear model

$$\mathbf{x}(n) = \mathbf{A}\mathbf{s}(n) + \mathbf{v}(n) \quad (1)$$

where  $\mathbf{A}$  is a mixture matrix (for example the array steering matrix in sensor array processing),  $\mathbf{s}(n)$  is a vector of signals, and  $\mathbf{v}(n)$  is noise from some possibly unknown distribution. Assuming the signal and noise are uncorrelated and zero-mean, the array covariance is

$$\mathbf{R}_x = E[\mathbf{x}(n)\mathbf{x}^H(n)] = \mathbf{A}\mathbf{R}_s\mathbf{A}^H + \mathbf{R}_v. \quad (2)$$

The problem considered is to determine the rank of the signal part/subspace, i.e.,  $d = \text{rank}(\mathbf{A}\mathbf{R}_s)$ , based on  $N$  observations of the data (1).

If the additive noise is spatially white,  $\mathbf{R}_v = \sigma^2\mathbf{I}$ , The (population) eigenvalues of (2) are

$$\lambda_1 \geq \dots \geq \lambda_d > \lambda_{d+1} = \dots = \lambda_m = \sigma^2, \quad (3)$$

i.e., the true noise eigenvalues are all equal. However, when calculated from the sample covariance

$$\hat{\mathbf{R}}_x = \frac{1}{N} \sum_{n=1}^N \mathbf{x}(n)\mathbf{x}^H(n), \quad (4)$$

estimated from a finite number of  $N$  data snapshots, the ordered sample eigenvalues are distinct with probability one, i.e.,

$$\hat{\lambda}_1 > \dots > \hat{\lambda}_d > \hat{\lambda}_{d+1} > \dots > \hat{\lambda}_m > 0. \quad (5)$$

The distribution  $F_{N\lambda}(\hat{\lambda})$  of (5) for a data sample of  $N$  snapshots, either in the form of a probability density function (PDF), or a cumulative distribution function (CDF),

tends to take a very complex form. The sample eigenvalues are biased (as in (5)) and mutually correlated. The exact distribution is only known for the Gaussian case with certain population eigenvalues, and is given in the form of a series expansion [8]. For the general case, both w.r.t. the actual source distribution and the population eigenvalues, the distribution (joint or marginals) may only be available asymptotically, for large  $N$  [1][8]. For small/moderate  $N$ , corresponding to many practical applications, the error in the asymptotics may be substantial. Thus, there is no general 'ease of use' form of  $F_{N\lambda}(\hat{\lambda})$  available.

Instead of relying on asymptotic results, which are unreliable on short data records, the detection scheme to be presented in the next section will be based on an approximate relation between the marginal distributions of the noise sample eigenvalues. Specifically, numerical experiments indicate that for the white noise case (3), the marginal PDFs of the noise sample eigenvalues are approximately related as

$$f_{N_i}(\hat{\lambda}_i) \cong f_N(\kappa^i \hat{\lambda}_i) \quad i \geq d+1 \quad (6)$$

for some  $\kappa$ , i.e., the marginals  $f_{N_i}(\hat{\lambda}_i)$ ,  $i \geq d+1$  are simply scaled versions of the same basic PDF  $f_N(\cdot)$ . While there is no claim of the generality of this approximation, it has shown to be very precise when the ratio  $N/m$  is say five or higher. Also, what is important for detection based on this property, is that the approximation is robust to slightly colored noise, and practically invariant to non-Gaussianity. Then, even if the data does not correspond perfectly to the assumed data model, (6) allows for robust rank detection. An example to illustrate (6) will be given in Section 4.

### 3. DETECTION

#### 3.1. Detection principle

To indicate how the relation (6) can be used for rank estimation, assume a number of  $m$  independent variables  $\eta_i$  having distributions identical to the marginal distributions of the sample eigenvalues  $\hat{\lambda}_i$ . From the  $\eta_i$ ,  $m-1$  secondary variables  $v_i$  are formed as the ratios

$$v_i = \eta_i / \eta_{i+1}, \quad i \in [1, m-1]. \quad (7)$$

Then, up to the order of the approximation given in (6),  $v_i$  for  $i \in [d+1, m-1]$  will have identical distributions, as these  $v_i$  are invariant to the (possibly unknown) scaling  $\kappa$ . However,  $v_d = \eta_d / \eta_{d+1}$ , involving the marginal of the smallest signal eigenvalue, will tend to larger values than  $v_{d+1}$ . This forms the basis for rank detection: if the marginals  $f_i(\lambda_i)$  can be captured from the data  $x(n)$ ,  $n \in [1, N]$ , the order  $d$  can be estimated by testing for equality among the distributions of  $v_i$ ,  $i \in [1, m-1]$ .

A practical algorithm to exploit this property for rank estimation is as follows:

1. Use the Bootstrap to first estimate the marginals of the sample eigenvalues  $f_{N_i}(\hat{\lambda}_i)$ ,  $i = [1, m]$ , and then the distributions of  $v_i$ ,  $i \in [1, m-1]$ .
2. Apply the Kolmogorov-Smirnov test [7] to test for pair-wise equality of the distributions  $F_{v,i}(v_i)$  of  $v_i$ , starting from the bottom ( $F_{v,m-2}(v_{m-2})$  versus  $F_{v,m-1}(v_{m-1})$ ), and stepping up until equality is rejected.

Before going into the full details of the scheme, it is necessary to establish how the Bootstrap behaves when resampling data to calculate eigenvalues.

#### 3.2. The Bootstrap and eigenvalues

The Bootstrap is a general tool for estimation of the distribution of a statistic from a sample of data. In this case the Bootstrap is employed to estimate  $F_{N\lambda}(\hat{\lambda})$ . The principle of the Bootstrap is as follows. The original data  $x(n)$ ,  $n = [1, N]$ , i.e.,

$$X_N = [x(1), \dots, x(N)], \quad (8)$$

is an estimate of the distribution of  $x(n)$ . Assigning each snapshot a probability  $1/N$ , resamples are taken randomly (with replacement) from  $X_N$ , giving Bootstrap data

$$X_N^* = [x^*(1), \dots, x^*(N)]. \quad (9)$$

From the Bootstrap resample  $X_N^*$ , the sample eigenvalues are calculated (through (4)), giving

$$\hat{\lambda}^* = [\hat{\lambda}_1^*, \dots, \hat{\lambda}_m^*] \quad (10)$$

with  $\hat{\lambda}_1^* > \dots > \hat{\lambda}_i^* > \dots > \hat{\lambda}_m^*$ . The procedure is repeated a number of  $B$  times. Then, the Bootstrap distribution derived from the  $B$  replicates of (10) is a nonparametric estimate of  $F_{N\lambda}(\hat{\lambda})$ .

As the sample eigenvalues are highly non-linear functions of the data sample, results on the Bootstrap w.r.t. linear statistics do not apply. Though, some results on the properties of eigenvalues calculated from resampled data can be found in [3][4]:

- For distinct population eigenvalues,  $\hat{F}_{N\lambda}(\hat{\lambda})$  converges asymptotically to  $F_{N\lambda}(\hat{\lambda})$ .
- For equal population eigenvalues (such as in the white noise case),  $\hat{F}_{N\lambda}(\hat{\lambda})$  does not converge to  $F_{N\lambda}(\hat{\lambda})$ . However, if resamples are taken of size  $M < N$  from  $X_N^*$ , such that  $M \rightarrow \infty$  as  $N \rightarrow \infty$ , while  $M/N \rightarrow 0$ , then  $\hat{F}_{M\lambda}(\hat{\lambda})$  converges weakly to  $F_{M\lambda}(\hat{\lambda})$ , i.e., the distribution of the eigenvalues of a sample  $x(n)$ ,  $n = [1, M]$ .

From numerical experiments it is easily seen that the major problem with the Bootstrap is to characterize the dependence between sample eigenvalues: while the Bootstrap does make a good job capturing the marginals, the dependence between the sample eigenvalues is not maintained in  $\hat{F}_{N\lambda}(\hat{\lambda})$  for reasonable  $N$ . This motivates the use of the

marginals only. Also, a full characterization of the joint  $m$ -dimensional distribution would require a very large data record ( $N$ ). By only considering the marginals, a much smaller data size is required. It is also worth considering resamples of size  $M < N$ . This relaxes the strong dependence on the actual data  $X_N$  somewhat, which seems to remove some erratic behavior seen on small sample sizes.

### 3.3. Detection scheme

The full estimation/detection procedure is as follows:

1. Estimate the marginal distributions  $f_{M_i}(\hat{\lambda}_i)$ ,  $i = [1, m]$ , by taking  $B$  resamples of size  $M$  from the data  $X_N$ . For each resample, calculate the sample eigenvalues  $\hat{\lambda}^*$  (10).
2. Estimate the distributions of  $v_i$ ,  $i \in [1, m-1]$  (7). To do this, note that in place of the fictitious *independent* variables  $\eta_i$ ,  $i \in [1, m]$ , sample eigenvalues  $\hat{\lambda}_i^*$  from *different* resamples  $\hat{\lambda}^*$  can be used (the sample eigenvalues from one resample are correlated). Thus, form

$$v_i^* = (\hat{\lambda}_i^*)_l / (\hat{\lambda}_{i+1}^*)_k, i \in [1, m-1] \quad (11)$$

with  $l$  and  $k$  being different resamples. Although an arbitrary number of resamples ( $B_2$ ) of (11) could be taken, it is sensible to use all  $B$   $\hat{\lambda}^*$  from step 1 in a systematic way. Estimate the CDFs of  $v_i$ ,  $i \in [1, m-1]$ , by the staircase approximation

$$\hat{F}_{v,i}(x) = \text{number of } (v_i^* < x) / B. \quad (12)$$

3. Determine the test statistics for the one-sided Kolmogorov-Smirnov (KS) test from the distributions (12)

$$T_i = \sup_x (\hat{F}_{v,i+1}(x) - \hat{F}_{v,i}(x)) \quad (13)$$

for  $i = [1, m-2]$ . Under the hypothesis that  $F_{v,i+1}(x)$  and  $F_{v,i}(x)$  are equal, the test statistic  $T_i$  is asymptotically distributed as [7]

$$P(\sqrt{B}T_i \leq x) \rightarrow 1 - \exp(-2x^2) \quad (14)$$

for  $x > 0$ .

4. Final step. Determine the rank  $d$  from a sequential test on the KS statistics:
  - I Set  $i = m-2$ .
  - II Define the null hypothesis  $H: d = i$ , and the alternative hypothesis  $K: d < i$ .
  - III Set a threshold  $\gamma$  based on the tail area of the distribution (14) of (13) under  $K$  [7].
  - IV If  $T_i > \gamma$  accept  $H$  (i.e. reject equality of distributions) and stop, else set  $i = i-1$  and return to II.

Note that in order to enable a correct decision, the test procedure requires there are at least two noise eigenvalues.

There are a number of parameters to be tuned/chosen in the scheme. First, consider the resample size  $M$ . A smaller  $M$  tends to improve the estimate of  $f_{M_i}(\hat{\lambda}_i)$ . How-

ever, a small  $M$  leads to a loss in the signal to noise ratio (SNR) detection threshold (i.e., the minimum SNR required for reliable rank detection), as the relative distance between  $f_{M_d}(\hat{\lambda}_d)$  and  $f_{M(d+1)}(\hat{\lambda}_{(d+1)})$  decreases with a decreasing  $M$ . For a data size  $N$  of order  $\sim O(10^2)$ , a reasonable trade-off is  $M \approx 3N/4$ .

The number of Bootstrap resamples  $B$  has an impact on the estimated distributions and is therefore a crucial parameter. Some guidelines on the impact of the number of Bootstraps  $B$  can be found in [5][6]. Unfortunately, no results are given in absolute terms. However, note that the proposed detection scheme does not require any critical values to be estimated with high precision. What is important is that the locations of the distributions of the  $v_i$  are estimated with sufficient accuracy for the subsequent KS test to work properly. Thus,  $B$  should be large enough such that the means of  $v_i^*$  are reasonably stable on a normalized scale. A coarse first order approximation of  $E[\hat{\mu}_{v_i}]$  gives

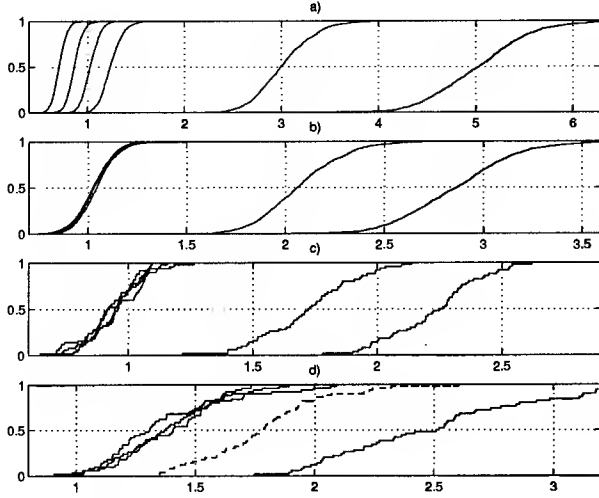
$$\hat{\mu}_{v_i} \approx \hat{\mu}_{\hat{\lambda}_i} (2/\mu_{\hat{\lambda}_{i-1}} - \hat{\mu}_{\hat{\lambda}_{i-1}}/\mu_{\hat{\lambda}_{i-1}}^2) \approx \hat{\mu}_{\hat{\lambda}_i}/\mu_{\hat{\lambda}_{i-1}}, \quad (15)$$

i.e., the stability of the location depends to a large extent on the location of  $\hat{\lambda}_i^*$ . To arrive at an expression for the necessary  $B$ , note that the sample eigenvalues are reasonably close to Gaussian. The separation (bias) of two sample eigenvalues corresponding to equal population eigenvalues is roughly two times the standard error, see Figure 1 (note that this relation holds regardless  $M$ ). Now, the standard error of the sample mean of  $B$  iid Gaussian variables is  $\sigma/\sqrt{B}$ . Thus, the location error of  $\hat{f}_{M_i}(\hat{\lambda}_i)$ , normalized to the separation of neighbouring distributions, is of order

$$\frac{\sigma/\sqrt{B}}{2\sigma} = \frac{1}{2\sqrt{B}}. \quad (16)$$

As an example, with  $B = 25$  the location error is of order 0.1 which is small enough for reliable detection. Note that there is no point using too large a  $B$ , as the error originating from the approximation (6) then will dominate the 'randomness' in  $T_i$ .

The final parameter to be chosen is the threshold  $\gamma$  for the KS test in Step 4. This threshold can be determined in two ways. First,  $\gamma$  can be set to maintain a desired level of the test at each sequential stage (as in the sphericity test), based on the distribution (14) of the test statistic (13) under  $K$  (the hypothesis that the distributions are equal). Alternatively,  $\gamma$  can be set for 'MDL-like consistency'. To see this, note that  $T_i \rightarrow 1$  rapidly under  $H$  for increasing SNR, or  $N$ . At the same time, under  $K$ , the tail probability of  $T_i$  is small even for modest  $\gamma$ . Thus,  $\gamma$  can be set to provide a probability of detection very close to one, without much penalty in the SNR threshold. As an example, with  $B = 25$ , the 95% level under  $K$  is  $\gamma \approx 0.35$ . With  $\gamma = 0.7$ , the level is 99.9995%.

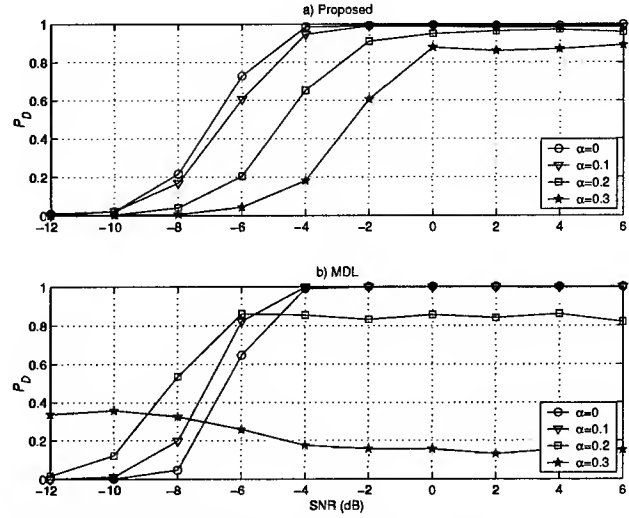


**Figure 1.** CDFs of a) sample eigenvalues, b) scaled sample eigenvalues, c) scaled Bootstrap eigenvalues, and d) the test variables  $(\lambda_i^*)_l / (\lambda_{i+1}^*)_k$ .

#### 4. NUMERICAL EXAMPLES

The detection scheme relies on the validity of the assumption (6). To illustrate the principle of the test, data was generated according to the model (1): a 6-element uniform linear array with half wavelength element spacing receives  $d = 2$  uncorrelated Gaussian signals from directions  $[10^\circ, 25^\circ]$ , relative to the array broadside. The signals were observed in white Gaussian noise with an element SNR of -3dB. Figure 1a shows the marginal CDFs of the 6 sample eigenvalues, when calculated based on  $N = 100$  independent array snapshots. Figure 1b shows the CDFs when the sample eigenvalues have been pre-scaled with  $\kappa^{i-4}$  (relative to eigenvalue number four) as in (6). In this case,  $\kappa \approx 1.21$ , and the scaled noise CDFs are all very close, with a largest pair-wise separation  $|F_{\kappa i} - F_{\kappa(i+1)}|$  of 0.11 for  $i > d$ .

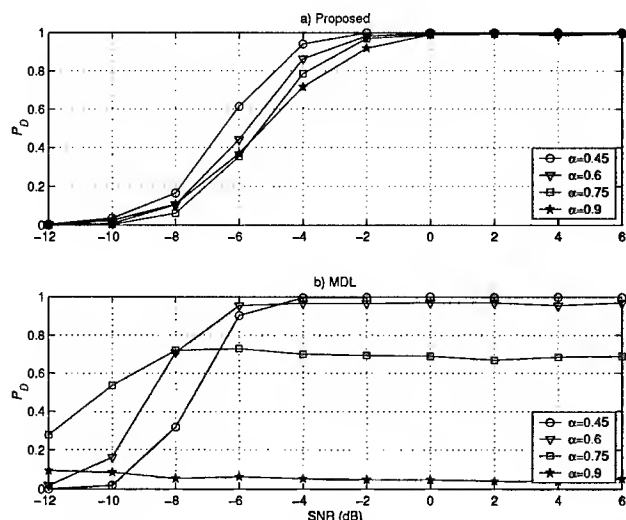
Similarly, Figure 1c shows the CDFs of scaled ( $\kappa \approx 1.36$ ) Bootstrap eigenvalues, estimated from  $B = 50$  resamples of size  $M = 75$ , taken from *one* data realization of  $N = 100$  snapshots. Clearly, the Bootstrap eigenvalues are slightly more variable, which is due both to  $M < N$ , and the effective loss in sample size from resampling. Again, the noise CDFs are close, but with some random fluctuations due to the limited number  $B$ . However, note that even with an infinite number of Bootstraps, there will still be a remaining error due to the approximation (6) (as in Figure 1b), as well as the limitation of the Bootstrap itself [3][4]. Finally, the CDFs of the variables (11), calculated from the  $B = 50$  sets of Bootstrap eigenvalues, are shown in Figure 1d. These are the CDFs on which the KS test is based. The 'noise only' CDFs are close, while the



**Figure 2.** The probability of correctly estimating the rank ( $d = 2$ ) versus the SNR, for various spatial noise color: a) Proposed scheme, b) MDL.

CDF of  $\hat{\lambda}_2 / \hat{\lambda}_3^*$  is the rightmost; with increasing SNR or data size this CDF moves further to the right. Clearly, the KS test can easily decide on the correct rank from the separation of the CDFs. Note that the dashed CDF is due to the two signal eigenvalues.

In the ideal case, with white Gaussian noise, the performance of the proposed scheme is virtually identical to MDL and the sphericity test (depending on how the threshold is set; for 'consistency', or for a fixed level) in terms of SNR and data size thresholds, and the ability to resolve closely spaced targets. Instead, the power of the new method lies in its robustness to unknown noise color. To illustrate, data was generated according to Figure 1, but varying the spatial noise color. Specifically, the  $k$ th element of the noise covariance matrix in (2) was  $(R_v)_{kl} = \exp(-\alpha|k-l|)$ , with  $\alpha$  being the parameter to be varied. For the detection procedure,  $B = 25$  resamples of size  $M = 75$  were taken from each original data set of size  $N = 100$ . The actual  $B$  is at a boundary: a smaller  $B$  leads to a penalty in SNR threshold, whereas a larger gives no further improvement. The threshold  $\gamma$  was set to 0.7 for 'consistent' detection. The performance of the proposed scheme as well as MDL as a function of the SNR is shown in Figure 2a-b, for  $\alpha \in [0, 0.1, 0.2, 0.3]$ . It is seen that the proposed scheme maintains good detection performance for increasing  $\alpha$ . Though, there is a penalty in the low SNR threshold. This is caused by the distributions of the noise eigenvalues being further separated for an increasing  $\alpha$ , leading to a reduction in the SNR margin. Increasing  $\alpha$  beyond 0.3 causes a more substantial degradation as the approximation (6) is no longer good. The performance of MDL suffers at comparatively small  $\alpha$ .



**Figure 3.** The probability of correctly estimating the rank ( $d = 2$ ) versus the SNR, for various temporal noise color: a) Proposed scheme, b) MDL.

Similarly to an unknown spatial noise correlation, an unknown temporal noise correlation may also lead to a degradation in detection performance. The above experiment was repeated with spatially white but temporally colored noise, having a temporal covariance of  $r(\tau) = \exp(-\alpha|\tau|)$ , with varying  $\alpha$ . With temporally colored noise, the data is no longer iid. For a better result with the Bootstrap, resampling was performed using block resampling [9]. The resample size was  $M = 70$ , with each resample made up of 7 random sections of 10 consecutive snapshots from the original data (with replacements). The number of resamples was increased to  $B = 50$ .

The results for various SNR and  $\alpha$  are shown in Figure 3a-b. As seen, MDL loses performance with an increasing  $\alpha$ . This is easily explained, as a temporal correlation reduces the effective data size. With the 'true' data size  $N$  being incorrect, the penalty term in MDL will be erroneous. On the other hand, the proposed technique does not rely directly on  $N$ , making it robust to the temporal noise correlation.

## 5. CONCLUSIONS

A new technique for rank estimation has been presented. While giving similar performance as classical well-known techniques under ideal conditions, the new method, based on the Bootstrap, is robust to errors in the noise model. The price for robustness is an increase in the computational complexity. However, as the number of Bootstrap replications is fairly small, this increase is modest.

## 6. REFERENCES

- [1] T. W. Andersson, 'Asymptotic Theory for Principal Component Analysis', *Ann. Math. Statist.*, vol. 34, 1963.
- [2] T. W. Andersson, 'An Introduction to Multivariate Statistical Analysis, 2nd ed.', Wiley, 1984.
- [3] R. Beran, M. S. Shrivastava, 'Bootstrap Tests and Confidence Regions for Functions of a Covariance Matrix', *Annals of Statistics*, vol. 13, no. 1, 1985.
- [4] R. Beran, M. S. Shrivastava, 'Correction-Bootstrap Tests and Confidence Regions for Functions of a Covariance Matrix', *Annals of Statistics*, vol. 15, no. 1, 1987.
- [5] B. Efron, B. Tibshirani, 'An Introduction to the Bootstrap', Chapman and Hall, 1993.
- [6] P. Hall, 'On the Number of Bootstrap Simulations Required to Construct a Confidence Interval', *Annals of Statistics*, vol. 14, no. 4, 1986.
- [7] E. B. Manoukian, 'Modern Concepts and Theorems of Mathematical Statistics', Springer, 1986.
- [8] R. J. Muirhead, 'Latent Roots and Matrix Variates: A Review of Some Asymptotic Results', *Annals of Statistics*, vol. 6, no. 1, 1978.
- [9] D. N. Politis, 'Computer-Intensive Methods in Statistical Analysis', *IEEE Signal Processing Magazine*, Jan. 1998.
- [10] B. Porat, 'Digital Processing of Random Signals', Prentice Hall, 1994.
- [11] P. Stoica, M. Cedervall, 'Detection Tests for Array Processing in Unknown Correlated Noise Fields', *IEEE Trans. Signal Processing*, vol. 45, no. 9, Sept. 1997.
- [12] Q. Wu, K. M. Wong, 'Determination of the Number of Signals in Unknown Noise Environments-PARADE', *IEEE Trans. Signal Processing*, vol. 43, no. 1, Jan. 1995.

# DETECTION-ESTIMATION OF MORE UNCORRELATED SOURCES THAN SENSORS IN NONINTEGER SPARSE LINEAR ANTENNA ARRAYS

Yuri I. Abramovich, Nicholas K. Spencer

Cooperative Research Centre for Sensor Signal and Information Processing (CSSIP),  
SPRI Building, Technology Park Adelaide, Mawson Lakes, South Australia, 5095, Australia

yuri@cssip.edu.au

nspencer@cssip.edu.au

## ABSTRACT

We introduce a new approach for the detection-estimation problem for sparse linear antenna arrays comprising  $M$  identical sensors whose positions may be noninteger values (expressed in half-wavelength units). This approach considers the (noninteger)  $M_\alpha$ -element co-array as the most appropriate *virtual array* to be used in connection with the augmented covariance matrix. Since the covariance matrix derived from such virtual arrays are usually very underspecified, we discuss a maximum-likelihood (ML) completion philosophy to fill in the missing elements of the partially specified Hermitian covariance matrix. Next, a transformation of the resulting unstructured ML matrix results in a sequence of properly structured positive-definite Hermitian matrices, each with their  $(M_\alpha - \mu)$  smallest eigenvalues being equal, appropriate for the candidate number of sources  $\mu$ . For each candidate model ( $\mu = 1, \dots, M_\alpha - 1$ ), we then find the set of directions-of-arrival (DOA's) and powers that yield the minimum fitting error for the specified covariance lags in the neighbourhood of the MUSIC-initialised DOA's. Finally, these models describe a hypothesis with respect to the actual number of sources, and allow us to select the "best" hypothesis using traditional information criteria (AIC, MDL, MAP, etc.) that are based on likelihood ratio.

## 1. INTRODUCTION

In our previous papers [5, 3, 2, 4], we introduced a new technique for detection-estimation of more uncorrelated Gaussian sources  $m$  than sensors  $M$  ( $m \geq M$ ) for the class of integer-spaced arrays. Here, we present one attempt to extend this approach to the class of noninteger-spaced nonuniform linear arrays (NLA's). Since such arrays generate up to  $\frac{1}{2}M(M-1)$  distinct nonzero covariance lags, they have the potential [8] to

estimate a *superior* number of uncorrelated Gaussian sources, *ie.* for the number of sources in the range

$$M \leq m \leq \frac{1}{2}M(M-1). \quad (1)$$

For a known number of sources  $m$ , we previously introduced [6] a DOA estimation technique capable of handling these superior scenarios. The current problem of detection-estimation is more complicated since we now require both an estimation of the number of sources and their DOA's.

Naturally, this problem has a solution if and only if the *identifiability* conditions hold, which in this case means that the observed set of covariance lags generated by the NLA can be uniquely decomposed into some number of signal dyads plus white noise. While the nonidentifiability conditions for detection are given in [4], here we concentrate on identifiable scenarios only; that is, for the true (deterministic) covariance lags and the chosen virtual array, the partially specified covariance matrix has a unique completion that corresponds to a mixture of  $m$  uncorrelated plane waves in white noise.

In practice, when the observed specified covariance lags are stochastic, being produced by a sample  $M$ -variate covariance matrix, the feasibility conditions for our type of positive-definite (p.d.) completion are not guaranteed. Therefore, in order to achieve a p.d. completion with equalised  $(M_\alpha - m)$  minimum eigenvalues, even the specified (measured) covariance lags need to be modified. Clearly, by not limiting the size of the modification of the specified lags, we can achieve a p.d. completion with the desired number  $(M_\alpha - \mu)$  of minimum eigenvalues being equal.

Note that for a Hermitian matrix to represent a mixture of  $\mu$  uncorrelated plane waves in noise, the equality of the  $(M_\alpha - \mu)$  smallest eigenvalues is only a necessary condition (whereas this is the necessary

and sufficient condition for a Toeplitz matrix). Thus some further modification of the specified covariance lags is required in order to correctly model the sources, along with an appropriate completion of the missing (unspecified) covariance lags.

In this way, we finally obtain a number of candidate models, *ie.*  $M_\alpha$ -variate p.d. Hermitian matrices of the proper structure, that are now compared with the ML completion discussed below using traditional information criteria that judge a loss in likelihood ratio against an overestimated number of sources.

## 2. PROBLEM FORMULATION

Consider  $m$  narrow-band plane-wave signals of power  $p \equiv [p_1, \dots, p_m]$  impinging upon a nonuniform linear array of  $M$  identical omnidirectional sensors located at positions  $\mathbf{d} \equiv [d_1 \equiv 0, d_2, \dots, d_M]$  measured in half-wavelength units. In the detection-estimation problem, the number of sources  $m$  is unknown. Adopting the commonly used data model [12], we have

$$\mathbf{y}(t) = S(\boldsymbol{\theta}) \mathbf{x}(t) + \boldsymbol{\eta}(t) \quad \text{for } t = 1, \dots, N \quad (2)$$

where

$$\mathbf{x}(t) = [x_1(t), \dots, x_m(t)]^T \quad (3)$$

$$\mathbf{y}(t) = [y_1(t), \dots, y_M(t)]^T \quad (4)$$

$$\boldsymbol{\eta}(t) = [\eta_1(t), \dots, \eta_M(t)]^T, \quad (5)$$

$x_j(t)$  ( $j = 1, \dots, m$ ) is the complex signal amplitude of the  $j^{\text{th}}$  plane wave, and where  $y_k(t)$  and  $\eta_k(t)$  ( $k = 1, \dots, M$ ) are the sensor output and the noise at the  $k^{\text{th}}$  sensor respectively. To permit DOA estimation in the superior case ( $m \geq M$ ), we restrict ourselves to the class of independent (Gaussian) signal amplitudes  $\mathbf{x}(t) \in \mathcal{C}^{m \times 1}$  such that

$$\mathcal{E}\{\mathbf{x}(t_1) \mathbf{x}^H(t_2)\} = \begin{cases} P \equiv \text{diag}[p] & \text{for } t_1 = t_2 \\ 0 & \text{for } t_1 \neq t_2, \end{cases} \quad (6)$$

We assume that the additive noise  $\boldsymbol{\eta}(t) \in \mathcal{C}^{M \times 1}$  is white and Gaussian:

$$\mathcal{E}\{\boldsymbol{\eta}(t_1) \boldsymbol{\eta}^H(t_2)\} = \begin{cases} p_0 I_M & \text{for } t_1 = t_2 \\ 0 & \text{for } t_1 \neq t_2. \end{cases} \quad (7)$$

The array manifold matrix is  $S(\boldsymbol{\theta}) \equiv [\mathbf{s}(\theta_1), \dots, \mathbf{s}(\theta_m)] \in \mathcal{C}^{M \times m}$ , where each constituent “steering vector”  $\mathbf{s}(\theta_j)$  is defined as

$$\mathbf{s}(\theta_j) = [1, \exp(i\pi d_2 w_j), \dots, \exp(i\pi d_M w_j)]^T \quad (8)$$

with  $w = \sin \theta \in [-1, 1]$ .

According to this model, the  $M$ -variate spatial covariance matrix

$$R = S P S^H + p_0 I_M \quad (9)$$

is p.d. Hermitian. Note that in our (“superior”) case of  $m \geq M$ , the noise-free covariance matrix  $S P S^H$  is generally of full rank. Given  $N$  independent samples (“snapshots”), the sufficient statistic for DOA estimation is the  $M$ -variate direct data covariance (DDC) matrix

$$\hat{R} = \frac{1}{N} \sum_{t=1}^N \mathbf{y}(t) \mathbf{y}^H(t). \quad (10)$$

To illustrate our technique, consider the “quasi-integer” [6] four-element NLA

$$\mathbf{d}_4 = [0, 1.09, 3.96, 5.93] \quad (11)$$

that may be easily recognised as a slightly perturbed version of the optimal four-element integer array [10]  $\mathbf{d} = [0, 1, 4, 6]$ . In [6], we demonstrated that up to six independent sources could be unambiguously identified by the NLA  $\mathbf{d}_4$ . The co-array of  $\mathbf{d}_4$  (the sorted set of nonduplicated position differences) is

$$\mathbf{c}_4 = [0, 1.09, 1.97, 2.87, 3.96, 4.84, 5.93] \quad (12)$$

and so the augmented  $M_\alpha = 7$ -variate Hermitian covariance matrix for the virtual array  $\mathbf{c}_4$  is extremely underspecified:

$$H = \begin{bmatrix} r_0 & r_{1.09} & r_{1.97} & r_{2.87} & r_{3.96} & r_{4.84} & r_{5.93} \\ r_{1.09}^* & r_0 & ? & ? & ? & ? & ? \\ r_{1.97}^* & ? & r_0 & ? & ? & ? & ? \\ r_{2.87}^* & ? & ? & r_0 & ? & ? & ? \\ r_{3.96}^* & ? & ? & ? & r_0 & ? & ? \\ r_{4.84}^* & ? & ? & ? & ? & r_0 & ? \\ r_{5.93}^* & ? & ? & ? & ? & ? & r_0 \end{bmatrix} \quad (13)$$

Nevertheless, it is important to understand that for the true covariance lags  $r_0, \dots, r_{5.93}$ , identifiability means that there exists a single p.d. completion of  $H$  with equalised  $(M_\alpha - \mu)$  minimum eigenvalues for any scenario with  $m < 6$  independent sources.

Let  $\mathcal{S}$  be the set of specified elements  $\{p, q\}$ , and  $\bar{\mathcal{S}}$  be the set of unspecified elements in the initial incomplete augmented covariance matrix  $H$ . Suppose for the moment that given the specified sample covariance lags  $r_{\mathcal{S}}$ , we somehow generate a set of candidate p.d.  $M_\alpha$ -variate Hermitian matrices  $H_\mu$  ( $\mu = 1, \dots, M_\alpha - 1$ ) that each correspond to the model of  $\mu$  plane waves in noise. To select the best candidate model, we calculate the likelihood ratio (LR) for each corresponding



$M$ -variate Hermitian matrix

$$R_\mu = L H_\mu L^T \quad (14)$$

where  $L$  is the  $M \times M_\alpha$  binary selection (or incidence) matrix with  $L_{jk}$  equal to unity in the  $j^{\text{th}}$  row and  $d_k^{\text{th}}$  column, and zero otherwise.

If we use the sphericity test [11]

$$\begin{aligned} H_0 : \mathcal{E} \left\{ R_\mu^{-\frac{1}{2}} \hat{R} R_\mu^{-\frac{1}{2}} \right\} &= p_0 I_M \quad \text{against} \\ H_1 : \mathcal{E} \left\{ R_\mu^{-\frac{1}{2}} \hat{R} R_\mu^{-\frac{1}{2}} \right\} &\neq p_0 I_M, \quad p_0 > 0 \end{aligned} \quad (15)$$

then the LR is

$$\gamma(H_\mu) = \frac{\det(R_\mu^{-1} \hat{R})}{\left[ \frac{1}{M} \text{Tr}(R_\mu^{-1} \hat{R}) \right]^M}. \quad (16)$$

Now information or Bayesian criteria may be used for model selection, such as the minimum description length [7]

$$\hat{m}_{MDL} = \arg \min_{\mu=0, \dots, M_\alpha-2} \left[ -\log \gamma(T_\mu) + \frac{3}{2} \mu \log N \right]. \quad (17)$$

Obviously this approach is optimal only if  $H_\mu$  is the ML estimate of the p.d. Hermitian matrix with equal  $(M_\alpha - \mu)$  minimum eigenvalues. Since exact ML estimates of this kind are not yet available, our problem is to generate a set of above-described Hermitian matrices  $H_\mu$  sufficiently close to the sufficient statistic  $\hat{R}$  in the ML sense.

### 3. "MAXIMUM-LIKELIHOOD" POSITIVE-DEFINITE HERMITIAN COMPLETION

In [6], we introduced several p.d. Hermitian completions, including maximum-entropy (ME) completion. These completions are used here as an initialisation step for the following optimisation routine. Let the general virtual array  $\mathbf{d}'$  be specified by the virtual sensor positions  $d'_j$  ( $j = 1, \dots, M_\alpha$ ), then the set of all possible p.d. Hermitian completions  $\mathcal{H}$  may be written as  $\mathcal{H} =$

$$\left\{ \mathbf{z} : H(\mathbf{z}) = H_0 + \sum_{\substack{p, q \in \bar{S} \\ p < q}} (\text{Re } H_{pq} E_+^{pq} + i \text{Im } H_{pq} E_-^{pq}) > 0 \right\} \quad (18)$$

where

$$\mathbf{z} = \begin{bmatrix} \text{Re } H_{pq} \\ \text{Im } H_{pq} \end{bmatrix}_{pq \in \bar{S}; p < q} \quad (19)$$

$$E_+^{pq} = \mathbf{e}_p \mathbf{e}_q^T + \mathbf{e}_q \mathbf{e}_p^T, \quad E_-^{pq} = \mathbf{e}_p \mathbf{e}_q^T - \mathbf{e}_q \mathbf{e}_p^T \quad (20)$$

$\mathbf{e}_p = [0, \dots, 0, 1, 0, \dots, 0]$  is the  $M_\alpha$ -variate basis vector with a unit entry in the  $p^{\text{th}}$  position, and  $H_0$  is the initial completion (eg. ME completion).

Suppose we label each of the missing lags ( $pq \in \bar{S}$ ;  $p < q$ ) from 1 to  $\ell$ , the total number of missing lags. For nonredundant NLA geometries such as  $\mathbf{d}_4$ , the number of missing lags is rather large:

$$\ell = \frac{1}{2}(\nu - 1)(\nu - 2), \quad (21)$$

where  $\nu = \frac{1}{2}M(M - 1) + 1$ . Now, instead of (18), we may write

$$\mathcal{H} = \left\{ \mathbf{z} : H(\mathbf{z}) = H_0 + \sum_{j=1}^{2\ell} z_j F_j > 0 \right\} \quad (22)$$

where

$$z_j = \begin{cases} \text{Re } r_{pq \in \bar{S}; p < q} & \text{for } j = 1, \dots, \ell \\ \text{Im } r_{pq \in \bar{S}; p < q} & \text{for } j = \ell + 1, \dots, 2\ell \end{cases} \quad (23)$$

$$F_j = \begin{cases} E_+^{pq} & \text{for } j = 1, \dots, \ell \\ i E_-^{pq} & \text{for } j = \ell + 1, \dots, 2\ell \end{cases} \quad (24)$$

For sufficiently small  $z_k$  in (22)

$$|z_k| \leq \varepsilon \quad \text{for } k = 1, \dots, 2\ell \quad (25)$$

we may treat the term  $\sum_{k=1}^{2\ell} z_k F_k$  as being equal to a perturbation matrix  $\delta H(\mathbf{z}_0)$ , and find a first-order expansion for the eigenvalues [9]. According to the sphericity LR (16), the problem of ML maximisation is associated with the problem of eigenvalue equalisation in the matrix

$$G(\mathbf{z}) \equiv \hat{R}^{-\frac{1}{2}} [L H(\mathbf{z}) L^H] \hat{R}^{-\frac{1}{2}}. \quad (26)$$

By applying a first-order expansion to the eigenvalues of  $G(\mathbf{z})$ :

$$G(\mathbf{z}) = G_0 + \sum_{k=1}^{2\ell} z_k \hat{R}^{-\frac{1}{2}} L F_k L^H \hat{R}^{-\frac{1}{2}} \quad (27)$$

we can derive that

$$\lambda_g[G(\mathbf{z})] = \lambda_g[G_0] + \sum_{k=1}^{2\ell} z_k \mathbf{u}_g^{(0)H} \hat{R}^{-\frac{1}{2}} L F_k L^H \hat{R}^{-\frac{1}{2}} \mathbf{u}_g^{(0)} \quad (28)$$

where  $\mathbf{u}_g^{(0)}$  ( $g = 1, \dots, M$ ) is the  $g^{\text{th}}$  eigenvector of the matrix  $G_0$ , with corresponding eigenvalue  $\lambda_g[G_0]$ . Now we can introduce the  $(M \times 2\ell)$  matrix

$$\mathcal{D}^{(0)} = \left\{ \mathbf{u}_g^{(0)H} \hat{R}^{-\frac{1}{2}} L F_k L^H \hat{R}^{-\frac{1}{2}} \mathbf{u}_g^{(0)} \right\}_{k=1, \dots, 2\ell}^{g=1, \dots, M} \quad (29)$$

and our search to find sufficiently small perturbations ( $|z_k| \leq \varepsilon$ ) that minimise the difference between the  $(M_\alpha - \mu)$  smallest eigenvalues of  $G(z)$  may then be formulated as the following linear programming (LP) problem:

$$\text{Find } \min (\alpha - \beta) \quad \text{subject to} \quad (30)$$

$$\lambda^{(0)} + \mathcal{D}^{(0)} z < \alpha \mathbf{1}, \quad \alpha > 0 \quad (31)$$

$$\lambda^{(0)} + \mathcal{D}^{(0)} z > \beta \mathbf{1}, \quad \beta > 0 \quad (32)$$

$$-\varepsilon < z_k < \varepsilon \quad \text{for } k = 1, \dots, 2\ell \quad (33)$$

where  $\lambda^{(0)}$  is the vector of noise-subspace eigenvalues:

$$\lambda^{(0)} = [\lambda_{M_\alpha - \mu + 1}^{(0)}, \dots, \lambda_{M_\alpha}^{(0)}]^T. \quad (34)$$

and  $\mathbf{1} \equiv [1, \dots, 1]^T$ . Let the solution of this LP problem be  $z^{(0)}$ , then we define an updated Hermitian matrix

$$H^{(1)} = H^{(0)} + \sum_{k=1}^{2\ell} z_k^{(0)} F_k \quad (35)$$

and so by direct decomposition

$$G^{(1)} = \hat{R}^{-\frac{1}{2}} L H^{(1)} L^H \hat{R}^{-\frac{1}{2}} \quad (36)$$

we may check the validity of the constraints (33), and decrease the perturbation "step size"  $\varepsilon$  if our equalisation step failed to improve the current differences amongst the noise-subspace eigenvalues of the matrix  $H^{(1)}$ . If the validity conditions are met, then we compute the associated  $u_g^{(1)}$  and  $\lambda^{(1)}$  and then solve the iterated LP problem. Suppose that  $\kappa$  iterations are required before this procedure essentially reaches its final stable point.

Naturally, the global optimality of the overall procedure cannot be guaranteed, whereas at each step (*ie.* locally), the LP routine provides the optimal solution.

Note that during this first stage of our routine, only the unspecified (missing) elements of  $H^{(\kappa)}$  have been varied, while the specified sample covariance lags remain the same as for the initial point  $H_0$ .

Now, during the second stage of the ML maximisation routine, we modify *all* covariance lags. Since small perturbations in the sample covariance lags of  $\hat{R}$  (with respect to the exact values in  $R$ ) lead to significant fluctuations in the noise-subspace eigenvalues  $\sigma_n$  of the matrix  $\hat{H}$ , "inverse perturbations" in  $\hat{H}$  that equalise up to the  $(M_\alpha - m)$  smallest eigenvalues should not involve significant changes to the sample covariance lags. Effectively, we use the same optimisation routine (30) here with the only significant difference that now all elements (except the diagonals) are varied, *ie.*

$$H^{(\kappa+1)} = H^{(\kappa)} + \sum_{k=1}^{M(M-1)} z_k^{(0)} F_k. \quad (37)$$

Given that we cannot guarantee the global optimality of this second optimisation routine also, we may treat the solution ( $H^{(\gamma)}$ , say) as the *unstructured* ML estimate of the  $M_\alpha$ -variate covariance matrix  $\hat{H}_{ML}$ . Therefore the probability of obtaining the desired number of identical minimum eigenvalues in  $\hat{H}_{ML}$  is zero.

For this reason, our third stage involves obtaining a properly structured ML estimate that corresponds to a mixture of  $\mu$  independent plane waves in noise. The unstructured ML estimate  $\hat{H}_{ML}$  is used as a sufficient statistic, and further modification of the unspecified entries occurs in order to equalise the  $(M_\alpha - \mu)$  smallest eigenvalues in this matrix. Obviously, we expect the more eigenvalues that are to be equalised, the more losses we will obtain in the LR compared with the ML estimate  $\hat{H}_{ML}$ .

Similarly to the above, we may present this equalisation routine as

$$H_\mu^{(j+1)} = H_\mu^{(j)} + \sum_{k=1}^{2\ell} z_k F_k, \quad H_\mu^{(0)} = \hat{H}_{ML} \quad (38)$$

where  $H_\mu^{(j)}$  is the p.d. Hermitian matrix obtained at the  $j^{th}$  iteration of the equalisation routine. As before, by applying a first-order perturbation expansion for the eigenvalues of the matrix  $H_\mu^{(j+1)}$ , we can derive the following LP problem:

$$\text{Find } \min (\alpha - \beta) \quad \text{subject to} \quad (39)$$

$$\sigma^{(j)} + \mathcal{V}^{(j)} z < \alpha \mathbf{1}, \quad \alpha > 0 \quad (40)$$

$$\sigma^{(j)} + \mathcal{V}^{(j)} z > \beta \mathbf{1}, \quad \beta > 0 \quad (41)$$

$$-\varepsilon < z_k < \varepsilon \quad \text{for } k = 1, \dots, 2\ell \quad (42)$$

where

$$\mathcal{V}^{(j)} = \left\{ \nu_i^{(j)H} F_k \nu_i^{(j)} \right\}_{k=1, \dots, 2\ell}^{i=M_\alpha - \mu + 1, \dots, M_\alpha} \quad (43)$$

$\nu_i^{(j)}$  is the  $i^{th}$  eigenvector of the matrix  $H_\mu^{(j)}$ , with associated eigenvalue  $\sigma_i^{(j)}$ , and  $\sigma^{(j)}$  is the vector of noise-subspace eigenvalues. Step size control of  $\varepsilon$  is implemented in the same fashion as before (33).

Clearly, the stable point of this third stage ( $H_\mu^{(J)}$ , say) would not result in exactly equal noise-subspace eigenvalues, since (as in the first stage) the specified entries have not been modified. Of course, it is possible to use a transformation to reach this final goal. Such a transformation keeps the eigenvectors of  $H_\mu^{(J)}$  invariant, and so the MUSIC-derived DOA estimates for  $\mu$  sources also remain the same. However, due to the dimension reduction brought about by (14), the

LR (16) would change as a result of such a transformation. Moreover, even with strictly equalised eigenvalues, the Hermitian matrix  $H_\mu^{(J)}$  does not necessarily correspond to the desired plane-waves-plus-noise model.

Thus our fourth and final stage, that considers the sequence of "ML" hypotheses  $H_\mu^{ML}$  ( $\mu = 1, \dots, M_\alpha - 1$ ), consists of a local ML refinement of the  $\mu$  DOA estimates and associated signal powers in the vicinity of the MUSIC DOA estimates generated by the covariance matrix  $H_\mu^{(J)}$ . This local refinement procedure is introduced in [1], and involves the specified covariance lags only. As a result, for each candidate model  $\mu = 1, \dots, M_\alpha - 1$ , we can find the "ML" set of estimated signal parameters  $\{\hat{\theta}_\mu^{ML}, \hat{p}_\mu^{ML}\}$  and estimated white noise power

$$\hat{p}_{0\mu}^{ML} = \frac{1}{M} \text{Tr} \hat{R} - \sum_{j=1}^{\mu} \hat{p}_{j\mu}^{ML} \quad (44)$$

that uniquely describes the covariance matrix  $R_\mu$  in the hypothesis (15)

$$R_\mu = \hat{p}_{0\mu}^{ML} I_M + \sum_{j=1}^{\mu} \hat{p}_{j\mu}^{ML} S(\hat{\theta}_\mu^{ML}) S^H(\hat{\theta}_\mu^{ML}). \quad (45)$$

Obviously, which ever information theoretic or Bayesian criterion is used for hypothesis selection, such a selection uniquely specifies not only the number of sources, but also the DOA and power estimates.

#### 4. FINAL COMMENTS

Simulation results (not introduced here) conducted for the NLA  $d_4$  for a superior number of sources demonstrates that the detection performance achieved by the four-stage algorithm described in this paper is comparable to that produced by the standard AIC and MDL criteria for conventional scenarios (with  $m < M$  sources) with the same Cramér–Rao bound. Naturally, in order to compare detection performance on conventional and superior scenarios, it is necessary to introduce significantly different intersource separation and/or sample sizes, however the comparable detection performance in the two cases suggests that the new detection scheme described here is close to optimum. An additional justification for this conclusion is that when our detection-estimation algorithm yields the true number of superior sources, we obtain a DOA estimation accuracy close to the corresponding Cramér–Rao bound.

#### REFERENCES

- [1] Y.I. Abramovich, D.A. Gray, A.Y. Gorokhov, and N.K. Spencer. Positive-definite Toeplitz completion in DOA estimation for nonuniform linear antenna arrays — Part I: Fully augmentable arrays. *IEEE Trans. Sig. Proc.*, 46 (9):2458–2471, 1998.
- [2] Y.I. Abramovich and N.K. Spencer. Detection-estimation of more uncorrelated Gaussian sources than sensors using partially augmentable sparse antenna arrays. In *Proc. EUSIPCO-2000*, Tampere, Finland. To appear September 2000.
- [3] Y.I. Abramovich, N.K. Spencer, and A.Y. Gorokhov. Detection of more uncorrelated Gaussian sources than sensors in nonuniform linear antenna arrays — Part I: Fully augmentable arrays. *IEEE Trans. Sig. Proc.* Submitted Feb 2000.
- [4] Y.I. Abramovich, N.K. Spencer, and A.Y. Gorokhov. Detection of more uncorrelated Gaussian sources than sensors in nonuniform linear antenna arrays — Part II: Partially augmentable arrays. *IEEE Trans. Sig. Proc.* In preparation.
- [5] Y.I. Abramovich, N.K. Spencer, and A.Y. Gorokhov. Detection of more uncorrelated Gaussian sources than sensors using fully augmentable sparse antenna arrays. In *Proc. SAM-2000*, Cambridge, MA, 2000.
- [6] Y.I. Abramovich, N.K. Spencer, and A.Y. Gorokhov. DOA estimation for noninteger linear antenna arrays with more uncorrelated sources than sensors. *IEEE Trans. Sig. Proc.*, 48 (4):943–955, 2000.
- [7] P.M. Djurić. A model selection rule for sinusoids in white Gaussian noise. *IEEE Trans. Sig. Proc.*, 44 (7):1744–1757, 1996.
- [8] J.-J. Fuchs. Extension of the Pisarenko method to sparse linear arrays. In *Proc. ICASSP-95*, pages 2100–2103, Detroit, 1995.
- [9] R.A. Horn and C.R. Johnson. *Matrix Analysis*. Cambridge University Press, England, 1990.
- [10] A.T. Moffet. Minimum-redundancy linear arrays. *IEEE Trans. Ant. Prop.*, 16 (2):172–175, 1968.
- [11] R.J. Muirhead. *Aspects of Multivariate Statistical Theory*. Wiley, New York, 1982.
- [12] P. Stoica and A. Nehorai. Performance study of conditional and unconditional direction-of-arrival estimation. *IEEE Trans. Acoust. Sp. Sig. Proc.*, 38 (10):1783–1795, 1990.

# A NEW GERSCHGORIN RADII BASED METHOD FOR SOURCE NUMBER DETECTION

*Hsien-Tsai Wu and Chan-Li Chen*

Department of Electronic Engineering,  
Southern Taiwan University of Technology  
No.1 Nan-Tai street, Yung Kung City, Tainan County, Taiwan

## ABSTRACT

In this paper, we introduce the effective uses of Gerschgorin radii [1-2] of the unitary transformed covariance matrix for source number detection. The heuristic approach applying a new Gerschgorin radii set developed from the projection concept, overcomes the problem in cases of small data samples and an unknown noise model. The proposed method is based on the sample correlation coefficient to normalize the signal Gerschgorin radii for source number detection. The performance of the proposed method shows improved detection capabilities over GDE [1,2] in Gaussian white noise process.

## 1. INTRODUCTION

Array processing, or more accurately, sensor array processing, is the processing of the output signals of an array of sensors located at different points in space in a wavefield. The purpose of array processing is to extract useful information from the received signals such as the number and location of the signal sources, the propagation velocity of waves, as well as the spectral properties of the signals. Array processing techniques have been employed in various areas in which very different wave phenomena occur. Common to all these applications, there are, in general, two essential purposes in array processing: (i) To determine the number of sources (decision), (ii) To estimate the locations of these sources (estimation).

Several high resolution detectors [3-5] for direction of arrival (DOA) have been developed in the field of passive underwater and radar signal processing in recent years. The primary contributions to the field include the MUSIC method proposed by Schmidt [3], the Minimum-Norm method by Kumaresan and Tufts [4], and the ESPRIT method by Roy et al. [5]. It is well known that the performances of these high resolution methods largely depend on the successful determination of the number of sources. Thus, several methods [6-11] have been suggested with this purpose in mind. Wax and Kailath [6] bring a statistical approach to solve the problem of source number detection based on the AIC and the MDL methods, which are generally used for the model selection.

In general, the AIC and MDL, including their modified versions, remain the most widely-used methods for estimating the source number. Most of them use the eigenvalues to estimate source number but neglect to use the eigenvectors as well. Consequently, Wu and Yang [1] proposed a heuristic approach by applying the Gerschgorin theorem to find Gerschgorin radii of the transformed covariance matrix for source number detection.

The heuristic detection criterion is developed from the concept of eigenvectors' projection.

In this paper, a proper similar transformation of the covariance matrix is required in order to effectively utilize the sample correlation coefficient to normalize the signal Gerschgorin radii for source number detection.

## 2. GERSCHGORIN DISK METHOD FOR SOURCE NUMBER DETECTION

### 2.1 Narrow-Band Model

We first review the narrow band mathematical model for estimating the number of sources and DOA of signals in a spatially white noise environment. The model we consider here consists of  $L$ -dimensional complex data vector  $\underline{x}(k)$  which represents the data received by an array of  $L$  sensors at the  $k$ th snapshot. The data vector is composed of plane-wave incident narrowband signals each of angular frequency  $\omega_0$  from  $M$  distinct sources embedded in Gaussian noise. Thus, the measured array data vector,  $\underline{x}(k)$ , which is assumed to be composed of  $M$  incoherent directional sources corrupted by additive white noise, is received at the  $k$ th snapshot by  $L$  ( $L > M$ ) sensors and is given by :

$$\underline{x}(k) = \sum_{i=1}^M s_i(k) \underline{a}(\omega_i) + \underline{n}(k) = \underline{A}(\omega) \underline{s}(k) + \underline{n}(k), \quad (1)$$

where  $\underline{A}(\omega) = [\underline{a}(\omega_1) \ \underline{a}(\omega_2) \ \dots \ \underline{a}(\omega_L)]$  is the direction matrix composed of direction vectors (steering vector) of the signals and the noise vector  $\underline{n}(k)$ , which is assumed to be complex, zero-mean, and Gaussian. The source vector is  $\underline{s}(k) = [s_1(k), s_2(k), \dots, s_M(k)]^T$ ,

where  $s_m(k)$  is the amplitude of the  $m$ th source and is assumed to be jointly circular Gaussian and independent of  $\underline{n}(k)$ . The exact form of the steering vector depends on the array configuration. However, the uniform linear array, apart from being most commonly used, may also offer advantageous implementation efficiency of some algorithms. For a propagation wavelength  $\eta$ , the distance between two sensors in a uniform linear array must be  $D \leq \eta/2$  and the corresponding steering matrix is given by

$$\underline{a}(\omega_m) = [1, \exp(j\omega_m), \dots, \exp(j(L-1)\omega_m)]^T, \quad (2)$$

where  $\omega_m$  is given by :  $\omega_m = 2\pi D \sin \theta_m / \eta$ , where  $D$  is the spacing between adjacent elements.  $\theta_m$  is the impinging angle of the  $m$ th source relative to the array broadside where  $\theta_m \in$

$(-\frac{\pi}{2}, \frac{\pi}{2})$  for all  $m$ . The vectors  $\underline{a}(\omega_m)$ ,  $m=1,2,\dots,M$  corresponding to  $M$  different values of  $\theta_m$  are assumed to be linearly independent. This implies that  $L > M$ , and  $\text{rank}(\underline{A})=M$ .

Note that it follows that  $\underline{x}(k)$  is a complex Gaussian vector with zero mean and covariance matrix given by

$$\underline{C} = E[\underline{x}(k) \underline{x}(k)^H] = \underline{A}(\omega) \underline{C}_s \underline{A}^H(\omega) + \sigma_n^2 \underline{I}, \quad (3)$$

where  $\underline{C}_s$ , which is the covariance matrix of  $\underline{s}(k)$ , is assumed to be non-singular, and  $\sigma_n^2$  is the variance of Gaussian noise. Superscripts  $*$ ,  $T$ , and  $H$  denote conjugate, transpose, and Hermitian transpose of matrices, respectively.

If  $N$  observations have been measured from  $L$  sensors, the entire data set can be placed in a  $L \times N$  matrix  $\underline{x}$  as:

$$\underline{x} = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1N} \\ x_{21} & x_{22} & \dots & x_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ x_{L1} & x_{L2} & \dots & x_{LN} \end{bmatrix} = [\underline{x}(1), \underline{x}(2), \dots, \underline{x}(k) \dots \underline{x}(N)]_{L \times N}. \quad (4)$$

Each row of  $\underline{x}$  represents a multivariate observation. For the  $L$ -dimensional scatterplot, the row of  $\underline{x}$  represent  $N$  points in  $L$ -dimensional space. Subsequently, the array sampled covariance matrix in Eq.(3) can also be expressed as :

$$\hat{\underline{C}} = \frac{1}{N} \underline{x} \underline{x}^H \quad (5)$$

## 2.2 Gerschgorin Disk Estimator

To make the Gerschgorin disk theorem effective, Wu et al. [2] proposed a proper transformation, called Gerschgorin Disk Estimator (GDE) for source number detection. The covariance matrix is first partitioned as :

$$\underline{C} = \begin{pmatrix} c_{11} & c_{12} & \dots & c_{1L} \\ c_{21} & c_{22} & \dots & c_{2L} \\ \vdots & \vdots & \ddots & \vdots \\ c_{L1} & c_{L2} & \dots & c_{LL} \end{pmatrix} = \begin{pmatrix} \underline{C}_1 & \underline{\epsilon} \\ \underline{\epsilon}^H & c_{LL} \end{pmatrix} \quad (6)$$

where  $\underline{C}_1$  is an  $(L-1) \times (L-1)$  leading principal submatrix of  $\underline{C}$ , which is obtained by deleting the last row and column of  $\underline{C}$ . Physically, it can be regarded as the removal of the  $L^{\text{th}}$  sensor. Thus,  $\underline{C}_1$  becomes the reduced covariance matrix of the remaining  $(L-1)$  sensors. The reduced covariance matrix  $\underline{C}_1$  also can be decomposed by its eigenstructure as :  $\underline{C}_1 = \underline{U}_1 \underline{D}_1 \underline{U}_1^H$ , where  $\underline{U}_1$  is an  $(L-1) \times (L-1)$  unitary matrix formed by the eigenvectors of  $\underline{C}_1$  as :

$\underline{U}_1 = [\underline{u}_1 \ \underline{u}_2 \ \dots \ \underline{u}_M \ \dots \ \underline{u}_{L-1}]$ , and  $\underline{D}_1$  is the diagonal matrix constructed from the corresponding eigenvalues as :

$$\underline{D}_1 = \text{diag}(\lambda'_1 \ \lambda'_2 \ \dots \ \lambda'_M \ \dots \ \lambda'_{L-1}). \quad (7)$$

The eigenvalues  $\lambda'_1 \geq \lambda'_2 \geq \dots \geq \lambda'_M \geq \lambda'_{M+1} \geq \dots \geq \lambda'_{L-1}$  are shown in descending order. Since  $\lambda'_i$  in Eq.(4) are the eigenvalues of the leading principal submatrix of  $\underline{C}$ , their eigenvalues satisfy the interlacing property shown as :  $\lambda_1 \geq \lambda'_1 \geq \lambda_2 \geq \dots \geq \lambda'_M \geq \lambda_{M+1} \geq \lambda'_{M+1} \geq \dots \geq \lambda_{L-1} \geq \lambda'_{L-1} \geq \lambda_L$ . The transformed covariance matrix becomes :

$$\underline{S} = \underline{U}^H \underline{C} \underline{U} = \begin{bmatrix} \lambda'_1 & 0 & 0 & 0 & 0 & 0 & \delta_1 \\ 0 & \lambda'_2 & 0 & 0 & 0 & 0 & \delta_2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \lambda'_M & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \lambda'_{L-1} & \delta_{L-1} \\ \delta_1^* & \delta_2^* & \dots & \delta_M^* & \dots & \delta_{L-1}^* & c_{LL} \end{bmatrix}, \quad (8)$$

where

$$\delta_i = \underline{u}_i^H \underline{\epsilon}, \quad (9)$$

for  $i=1, 2, \dots, L-1$ .

It is clear that the first  $(L-1)$  Gerschgorin disks ( i.e.  $\underline{O}_1, \underline{O}_2, \dots, \underline{O}_{L-1}$  ) possess the Gerschgorin radii :

$$r_i = |\rho_i| = |\underline{u}_i^H \underline{\epsilon}|, \quad (10)$$

for  $i = 1, 2, \dots, L-1$ . It is necessary to verify that all of the  $\rho_i$  values are equal to zero when  $i=(M+1), (M+2), \dots, (L-1)$  due to the fact that the noise eigenvectors,  $\underline{u}_i$ , are orthogonal to  $\underline{A}_1$ , which is the direction matrix of  $\underline{C}_1$ .

Since  $\underline{S}$  is a unitary transformation matrix of  $\underline{C}$ , they will share the same eigenvalues. The collection of the first  $(L-1)$  Gerschgorin disks,  $\underline{O}_i$ , contains its Gerschgorin center at  $c_i = \lambda'_i$  and the corresponding Gerschgorin radius  $r_i = |\rho_i|$ ,  $i = 1, 2, \dots, (L-1)$ . The disks with zero radii ( i.e.  $\underline{O}_{M+1}, \underline{O}_{M+2}, \dots, \underline{O}_{L-1}$  ) are regarded as the collection of noise Gerschgorin disks. The remaining disks ( i.e.  $\underline{O}_1, \underline{O}_2, \dots, \underline{O}_M$  ) containing non-zero radii and large centers are considered to be the source Gerschgorin disks. Hence, we can determine the number of sources by counting the number of non-zero Gerschgorin radii in the case of infinite samples. In addition, we can also use  $(L-1)$  eigenvalues of  $\underline{C}_1$  to determine the number of sources.

It can be seen that the threshold must be adjustable to varying numbers of snapshots. Hence, we define a heuristic decision rule

$$\text{as [2]: } GDE(k) = r_k - \frac{D(N)}{L-1} \sum_{i=1}^{L-1} r_i, \quad (11)$$

Where  $k$  is an integer in the closed interval  $[1, L-2]$ . The adjustable factor,  $D(N)$ , could be a non-increasing function (between 1 and 0) when  $N$  increases. If  $GDE(k)$  is evaluated from  $k=1$ , the number of sources is determined as  $k-1$  (i.e.  $M=k-1$ ) when the first nonpositive value of  $GDE(k)$  is reached. This is

due to the fact that the radius value below the adjustable threshold will be considered the noise collection. Thus, the above GDE rule may produce problems of underestimation

### 3. A NEW GERSCHGORIN RADII BASE METHOD

Hence, the method capable of reducing the radii size of signal Gerschgorin disks should help resolve source number detection problem.

#### 3.1 Correlation Coefficients of Samples Space

In light of these requirements, an effective source number detection method must select a proper transformation for maximum reduction of the radii size of signal Gerschgorin disks and make noise Gerschgorin disks as remote as possible from signal Gerschgorin disks. Therefore, a nonsingular matrix,  $\underline{D} = \text{diag}(\lambda'_1, \lambda'_2, \dots, \lambda'_M, \dots, \lambda'_{L-1}, 1)$  was used in [2] to get small signal

Gerschgorin radii, such as  $r'_i = \frac{r_i}{\lambda'_i}$ ,  $i=1,2,\dots,M$ . That method

led to development a novel technique, which outperformed GDE in Gaussian white and nonwhite noise processes and could be used successfully even when SNR is near 0 dB. In this paper, we extend the function of reducing signal Gerschgorin disks by using a new developed similar transformation of the sampled covariance matrix and its new set of normalized radii of signal Gerschgorin disks.

As Eq.(4), If  $N$  observations have been measured from  $L$  sensors, the entire data set can be also placed in a  $L \times N$  matrix  $\underline{x}$  as:

$\underline{x} = [\underline{x}(1), \underline{x}(2), \dots, \underline{x}(L-1), \underline{x}(L)]_{L \times N}$ . According to the definition of the multiple linear regression [12], the maximum correlation coefficient is define as

$$\rho_{ik} = \frac{|c_{ik}|}{\sqrt{c_{ii}}\sqrt{c_{kk}}} \quad (12)$$

for  $i=1, 2, \dots, L$  and  $k=1, 2, \dots, L$ . Note  $\rho_{ik} = \rho_{ki}$  for all  $i$  and  $k$ . The value of  $\rho_{ik}$  must be between 0 and +1.

Without altering the true eigenvalues, a proper transformation of the covariance matrix is required in order to effectively utilize the sample correlation coefficient to normalize the signal Gerschgorin radii for source number detection.

#### 3.2 The Proposed Method

In this section, a new transformation kernel based on the concept of sample correlation coefficient is proposed in order to improve detection performance. Now, a novel transforming matrix is proposed:

$$\begin{aligned} \underline{D} &= \text{diag}(\sqrt{c_{LL}}, \sqrt{\lambda'_1}, \dots, \sqrt{c_{LL}}, \sqrt{\lambda'_{L-1}}, 1) \\ &= \text{diag}(\Psi_1, \Psi_2, \dots, \Psi_M, \dots, \Psi_{L-1}, 1), \end{aligned} \quad (13)$$

to the transformed matrix in Eq.(14), where  $\lambda'_i$  are the eigenvalues of the first  $(L-1) \times (L-1)$  leading principal submatrix of  $\underline{C}$ .

The new transformed true covariance matrix becomes:

$$\underline{S}' = \underline{D}^{-1} \underline{U}^H \underline{C} \underline{U} \underline{D} = \underline{D}^{-1} \begin{pmatrix} \underline{U}_1^H \underline{C}_1 \underline{U}_1 & \underline{U}_1^H \underline{C}_2 \\ \underline{C}_2^H \underline{U}_1 & c_{LL} \end{pmatrix} \underline{D} \quad (14)$$

According to the Gerschgorin disk theorem, it is clear that the first  $(L-1)$  Gerschgorin disks (i.e.  $\mathbf{O}_1, \mathbf{O}_2, \dots, \mathbf{O}_{L-1}$ ) contain the new Gerschgorin radii :

$$r'_i = \frac{|\delta_i|}{\sqrt{\lambda'_i} \sqrt{c_{LL}}} = \frac{r_i}{\Psi_i}, \quad (15)$$

for  $i=1, 2, \dots, M$ . Since  $r'_i$  in Eq.(15) can be considered as the correlation coefficient of the covariance matrix in Eq.(12). The values of  $r'_i$  are all less than 1, so that  $\Psi_i > r_i$ . In other words, the disk size of signal Gerschgorin disks can be reduced as small as possible and the noise Gerschgorin disks can be kept as remote from the signal Gerschgorin disks as possible. Therefore, the source number can be easily determined by visually counting the number of signal Gerschgorin disks derived by Eq.(14). Moreover, when the noise statistics can not be accurately estimated, the GDE method fails under a low SNR situation; whereas the proposed method may not.

For example, in the case of one simulated covariance matrix, the sensor number is 6 (i.e.  $L=6$ ) and two sources (i.e.  $M=2$ ) are uncorrelated and impinged from  $-12^\circ$  and  $10^\circ$  (i.e.  $\text{DOA} = [-12^\circ, 10^\circ]$ ). The signal-to-noise ratios are both 2 dB (i.e.  $\text{SNR} = [10, 10]$  dB) and the number of samples chosen is  $N=100$ . Its Gerschgorin disks in terms of Gerschgorin center-and-radius pairs become  $\{12.11, 0.42\}$ ,  $\{7.93, 4.71\}$ ,  $\{0.19, 0.18\}$ ,  $\{0.09, 0.36\}$ , and  $\{0.08, 0.03\}$ . The results are illustrated in Figure 1(a). Subsequently, the same covariance matrix is transformed by the suggested unitary transformation as shown in Eq.(14). The results are illustrated in Figure 1(b). It is now significant that the Gerschgorin disks form two separate collections. The source collection contains disks  $\mathbf{O}_1$  and  $\mathbf{O}_2$  with small radii (less than 1) and the noise collection  $\mathbf{O}_3 \cap \mathbf{O}_4 \cap \mathbf{O}_5$  with small radii.

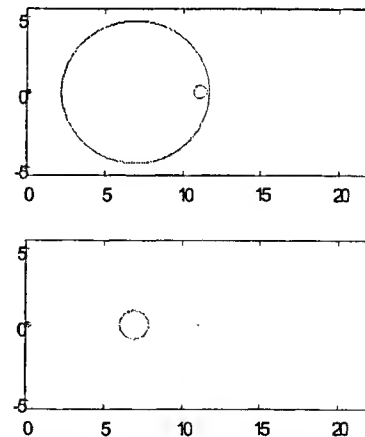


Fig.1(a)(b). Gerschgorin disks of the estimated covariance matrix

## 4. SIMULATION RESULTS

A uniformly linear array of 8 isotropic sensors is spaced a half wavelength apart with additive and uncorrelated white noise. The VGD and GDE methods are used to detect two uncorrelated sources with SNR's of 6dB impinging from  $0^\circ$  and  $5^\circ$  respectively. After 200 Monte Carlo runs, we compute their relative frequency of false detection using various numbers of snapshots. Error detection performance in terms of probabilities is depicted in Figure 2. It can be seen that the proposed method outperforms GDE.

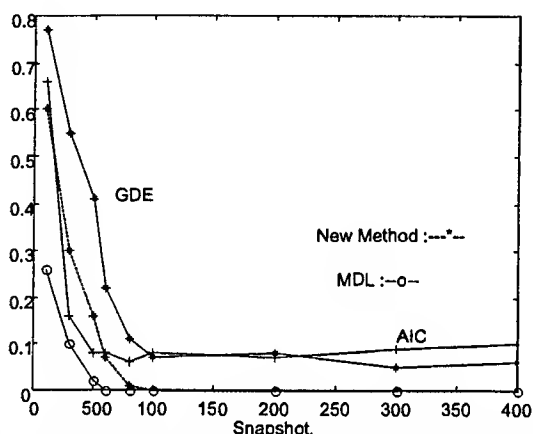


Fig.2 Detection performance of the AIC, MDL, GDE, and the proposed method in uses of simulated data with Gaussian white noise.(SNR=[6 6]dB, DOA=[ $0^\circ$   $5^\circ$ ])

## 5. CONCLUSION

In this paper, GDE performance is improved by using a developed similar transformation of the covariance matrix and using its new set of Gerschgorin radii to design the source number estimators. The proposed method is based on the sample correlation coefficient to normalize the signal Gerschgorin radii for source number detection. The performance of the proposed method shows detection capabilities superior to GDE in Gaussian white noise process and can be used successfully in a situation of measured experimental data.

## ACKNOWLEDGMENT

This research was supported by the National Science Council under Grant #NSC88-2612-E-218-001, Taiwan, Republic of China.

## 6. REFERENCES

- [1] H. T. Wu, J. F. Yang, and F. K. Chen, "Source number estimators using Transformed Gerschgorin Radii," *IEEE Trans. SP*, vol.43, pp.1325-1333, Jun. 1995
- [2] H. T. Wu, and J. F. Yang, "Gerschgorin radii based source number detection for closely spaced signals," *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, Atlanta, pp.3054-3057, May, 1996.
- [3] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," in *Proc. RADC Spectrum Estimation Workshop*, pp.243-258, Oct. 1979.
- [4] R. Kumaresan and D. W. Tufts, "Estimating the angles of arrivals of multiple plane waves," *IEEE Trans. Aerospace and Electronic Systems*, vol. AES-19, pp.134-139, 1983.
- [5] R. Roy and T. Kailath, "ESPRIT-estimation of signal parameters via rotational invariance techniques," *IEEE Trans. ASSP*, vol.ASSP-37, pp.984-995, July 1989.
- [6] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Trans. ASSP*, vol.33, no.2, pp.387-392, April 1985.
- [7] K. M. Wong, Q. T. Zhang, J. P. Reilly, and P. C. Yip, "On information theoretic criteria for determining the number of signals in high resolution array processing," *IEEE Trans. ASSP*, vol.38, no. 11, pp.1959-1970, Nov. 1990.
- [8] M. Wax, "Detection and localization of multiple sources in spatially colored noise," *IEEE Trans. SP*, vol.40, no.1, pp.245-249, Jan. 1992.
- [9] Q. Wu and D. R. Fuhrmann, "A parametric method for determining the number of signals in narrow-band direction finding," *IEEE Trans. SP*, vol.39, no.8, pp.1848-1857, Aug. 1991.
- [10] M. Wax, "Detection and localization of multiple sources in spatially colored noise," *IEEE Trans. SP*, vol.40, no.1, pp.245-249, Jan. 1992
- [11] W. Wu, J. Pierre, and M. Kaveh, "Practical detection with calibrated arrays," *Proc. of Statistical Signal and Array Processing Workshop*, pp.82-85, Canada, Oct. 1992.
- [12] Richard A. Johnson and Dean W. Wichern, *Applied Multivariate Statistical Analysis*, Prentice-Hall, Inc., New Jersey, 1988.

# ADAPTING MULTITAPER SPECTROGRAMS TO LOCAL FREQUENCY MODULATION

James W. Pitton

Applied Physics Laboratory, University of Washington  
1013 NE 40th St.  
Seattle, WA 98105  
E-mail: pitton@apl.washington.edu

## ABSTRACT

This paper presents further extensions to the multitaper time-frequency spectrum estimation method developed by the author. The method uses time-frequency (TF) concentrated basis functions which diagonalize the nonstationary spectrum generating operator over a finite region of the TF plane. Individual spectrograms computed with these eigenfunctions form direct TF spectrum estimates, and are combined to form the multitaper TF spectrum estimate. A method is presented for adapting the multitaper spectrogram to locally match frequency modulation in the signal, which can cause broadening of the spectral estimate. An F-test for detecting and removing frequency-modulated tones is also given.

## 1. INTRODUCTION

Thomson's multitaper spectral estimation approach [1] is a powerful method for nonparametric spectral estimation. This method uses a set of orthogonal data tapers that are maximally concentrated in frequency and diagonalize the spectral generating operator. These tapers are used to approximately invert the operator and estimate the spectrum. The multitaper approach was first applied to time-frequency (TF) analysis by a direct extension to the nonstationary case through a sliding-window framework [2], in which spectrograms are computed with each of the tapers and combined to form an estimate of the TF spectrum. A multitaper TF spectrum was constructed using spectrograms computed with Hermite windows [3], which had previously been shown to maximize a TF concentration measure [4]. This method was extended to include a means of reducing artifacts using a TF mask [5]. More recently, a multitaper method for TF analysis was pre-

sented by this author [6] that diagonalized the *nonstationary* spectral generating operator, formally extending Thomson's approach to TF. Subsequent work by the author gave bias and variance measures for the estimated TF spectrum, presented an adaptive procedure to reduce the bias of the individual spectrograms, and derived other properties of the eigenfunctions and the resulting TF spectral estimate [7, 8].

In this paper, a method is presented for adapting the multitaper spectrogram to locally match frequency modulation in the signal, which can cause broadening of the spectral estimate. Frequency modulation (FM) in the signal will degrade the resolution and accuracy of the multitaper spectrogram due to well-known spectral broadening effects. One common way of alleviating the effects of the spectral broadening is to match the spectrogram to the FM by frequency-modulating the window. This approach works perfectly well when there is only one FM rate in the signal, as is the case with chirped sonar and radar. However, in multicomponent signals such as speech, biological, and mechanical signals, there can be multiple FM rates present at any given time. To accurately analyze these types of signals, it is necessary to *locally* adapt the multitaper spectrogram to the FM at a given TF region. This paper presents a method for performing this local adaptation. An F-test for detecting and removing frequency-modulated tones is also given.

## 2. BACKGROUND: MULTITAPER TIME-FREQUENCY SPECTROGRAMS

This approach to TF spectral estimation is based on a straightforward extension of the spectral representation theorem for stationary processes [9], and is equivalent to a linear time-varying (LTV) filter model. Define the signal  $s(t)$  as the output of a white-noise-driven LTV

---

This work was supported by the National Science Foundation and the Office of Naval Research.



filter. The signal can then be written as:

$$s(t) = \int H(t, \omega) e^{j\omega t} dZ(\omega), \quad (1)$$

where  $H(t, \omega)$  is defined as the Fourier transform of the LTV filter  $h(t, t - \tau)$  [10]. The TF spectrum is defined by:

$$P(t, \omega) = |H(t, \omega)|^2. \quad (2)$$

This formulation for a TF spectrum is of the same general form as Priestley's evolutionary spectrum [9]; however,  $H(t, \omega)$  is not constrained to be slowly-varying.

Given a signal  $s(t)$ , an estimate  $P(t, \omega)$  is desired; however, direct inversion of equation (1) is impossible. A rough estimate of the time-varying frequency content of  $s(t)$  may be obtained by computing its short-time Fourier transform (STFT):

$$S_s(t, \omega) = \int s(\tau) g(t - \tau) e^{-j\omega \tau} d\tau, \quad (3)$$

where  $g(t)$  is a rectangular window of length  $T$ . A relationship between the STFT and  $H(t, \omega)$  is obtained by replacing  $s(t)$  by its TF spectral formulation:

$$S_s(t, \omega) = \int \int H(\tau, \theta) g(t - \tau) e^{-j(\omega - \theta)\tau} dZ(\theta) d\tau. \quad (4)$$

To solve for the time-varying spectrum  $H(\tau, \theta)$ , the STFT operator  $g(t - \tau) e^{-j\omega \tau}$  must be inverted. This inversion is an inherently ill-posed problem. Instead, the inverse solution is approximated by regularizing it to some region  $R(t, \omega)$  in the TF plane, much as Thomson regularized the spectral inversion to a bandwidth  $W$  in his multitaper approach [1]. For simplicity throughout,  $R(t, \omega)$  is defined to be a square TF region of dimension  $\Delta T \times \Delta W$ ; however, the results readily generalize to arbitrary regions.

In the case of spectral estimation, the operator is square and Toeplitz; its regularized inverse is found through an eigenvector decomposition. Such is not the case in the TF problem; the STFT operator is neither full rank nor square. This operator is diagonalized using a Singular Value Decomposition, giving left and right eigenvectors  $u(\tau)$  and  $V(t, \omega)$  and the associated eigen (singular) values  $\lambda$ :

$$g(t - \tau) e^{-j\omega \tau} = \sum_k \lambda_k u_k(\tau) V_k^*(t, \omega). \quad (5)$$

The eigenvectors  $u(\tau)$  and  $V(t, \omega)$  form an STFT pair:

$$V(t, \omega) = \int u(\tau) g(t - \tau) e^{-j\omega \tau} d\tau. \quad (6)$$

The SVD relationship between  $u(\tau)$  and  $V(t, \omega)$  is obtained by applying the STFT operator to  $V(t, \omega)$ , computing the integrals only over  $\Delta T \times \Delta W$ :

$$\lambda u(\tau) = \int_{\Delta T} \int_{\Delta W} V(t, \omega) g(t - \tau) e^{j\omega \tau} d\omega dt. \quad (7)$$

The inverse STFT computed over all  $(t, \omega)$  also holds. This equation can be reduced to a standard eigenvector equation by substituting for  $V(t, \omega)$ . The eigenvalue equation for  $u(\tau)$  is then:

$$\lambda u(\tau) = \int 2\Delta W \text{sinc}(\Delta W(\tau - s)) f(\tau, s) u(s) ds, \quad (8)$$

where

$$f(\tau, s) = \int_{\Delta T} g(t - s) g(t - \tau) dt. \quad (9)$$

$u(\tau)$  can be computed using standard eigenvalue solution methods. As has been discussed elsewhere, the eigenvectors are concentrated in TF and doubly orthogonal, both over the entire TF plane and over  $\Delta T \times \Delta W$ . These properties are critical for the estimation method.

Next,  $H(t, \omega)$  is estimated regularized to the rectangular region  $\Delta T \times \Delta W$  by projecting it onto  $\Delta T \times \Delta W$  in the vicinity of  $(t, \omega)$  using the  $k^{\text{th}}$  left eigenvector  $u_k(t)$ :

$$H_k(t, \omega) \doteq \lambda_k^{-\frac{1}{2}} \int_{\Delta T} \int_{\Delta W} H(\tau, \theta) u_k(t - \tau) e^{j(\theta - \omega)\tau} dZ(\theta) d\tau. \quad (10)$$

$H_k$  is thus a direct, but unobservable, projection of  $H(t, \omega)$  onto  $\Delta T \times \Delta W$ .

These expansion coefficients are then estimated using the STFT of  $s(t)$  computed using  $u_k(t)$ :

$$S_k(t, \omega) = \int \int H(\tau, \theta) u_k(t - \tau) e^{-j(\omega - \theta)\tau} dZ(\theta) d\tau, \quad (11)$$

i.e., the  $k^{\text{th}}$  eigenspectrum  $S_k(t, \omega)$  is a projection of  $H(t, \omega)$  onto the  $k^{\text{th}}$  left eigenvector  $u_k(t)$ , estimating  $H_k(t, \omega)$  over  $\Delta T \times \Delta W$ . When  $s(t)$  is a stationary white noise process, it follows that

$$E[|S_k(t, \omega)|^2] = |H(t, \omega)|^2 = P(t, \omega). \quad (12)$$

Thus, the individual eigenspectra are direct estimates of  $P(t, \omega)$ , and are unbiased when the spectrum is white.

Next,  $H(t, \omega)$  is estimated over  $\Delta T \times \Delta W$  using the right eigenvectors  $V_k(t, \omega)$  weighted by the projections of  $H(t, \omega)$  onto  $u_k(t)$ , i.e., the  $k^{\text{th}}$  spectrogram:

$$\hat{H}(\bar{t}, \bar{\omega}; t, \omega) = \sum_{k=1}^K V_k(\bar{t} - t, \bar{\omega} - \omega) S_k(t, \omega), \quad (13)$$

where  $K \approx \Delta T \Delta W$ . Choosing  $\Delta T \Delta W$  too small will result in estimates with poor bias and variance properties. The magnitude-square of  $\hat{H}(\bar{t}, \bar{\omega}; t, \omega)$  is an estimate of  $P(t, \omega)$  over  $\Delta T \times \Delta W$ . This estimate is a  $\chi^2$  random variable with two degrees of freedom (except for DC and Nyquist) with variance  $P^2(t, \omega)$ . The variance of this estimate can be reduced by averaging over  $\Delta T \times \Delta W$  and invoking the orthogonality of  $V_k(t, \omega)$ :

$$\begin{aligned}\hat{P}(t, \omega) &= \frac{1}{\Delta T \Delta W} \int_{\Delta T} \int_{\Delta W} |\hat{H}(\bar{t}, \bar{\omega}; t, \omega)|^2 d\bar{t} d\bar{\omega} \\ &= \frac{1}{\Delta T \Delta W} \sum_{k=1}^K \lambda_k |S_k(t, \omega)|^2.\end{aligned}\quad (14)$$

The average of  $K$  direct estimates is a  $\chi^2$  random variable with  $2K$  degrees of freedom; hence, the variance of this estimate is  $P^2(t, \omega)/K$ . If  $\Delta T$  is chosen to be a fixed proportion of the window length  $T$ , then this estimator is consistent for fixed  $\Delta W$ . Note that the form of this estimator differs slightly from that presented previously [6, 7, 8] in the weighting by the eigenvalues.

### 3. LOCALLY STATIONARY PROCESSES

The estimate for  $P(t, \omega)$  given in equation (14) is unbiased for white noise. For the estimate to be unbiased for signals other than white noise, it is only necessary that  $P(t, \omega)$  be *locally* white in TF, since the estimate is regularized to  $\Delta T \times \Delta W$ . A similar requirement is seen in the stationary case [1], wherein the spectrum is assumed to be smoothly varying so that it is approximately white over  $\Delta W$ . A class of stochastic processes known as *locally stationary* processes [12] satisfy the requirement of being smoothly varying in TF, and can be used to describe a wide variety of nonstationary signals. Locally stationary processes are stochastic processes with covariance functions of the form

$$R(t_1, t_2) = E[s(t_1)s^*(t_2)] = g\left(\frac{t_1 + t_2}{2}\right)f(t_1 - t_2), \quad (15)$$

where  $g(\cdot)$  is a nonnegative function and  $f(\cdot)$  is a valid covariance function; that is,  $f(t)$  possesses a nonnegative Fourier transform  $F(\omega)$ . Through a change of variables, the symmetric form of the covariance function is seen to be:

$$R_s(t, \tau) = E[s(t + \tau/2)s^*(t - \tau/2)] = g(t)f(\tau), \quad (16)$$

The TF spectrum is thus given by [11]:

$$P_s(t, \omega) = g(t)F(\omega). \quad (17)$$

For locally stationary  $s(t)$ ,  $P_s(t, \omega)$  will be approximately constant over  $\Delta T \times \Delta W$ , and equation (12) will still hold.

The class of processes with such nonnegative TF spectra is easily extended to include a wider range of nonstationary processes [13]. Let  $s(t)$  be a locally stationary process with covariance function  $R_s(t, \tau)$  and corresponding TF spectrum  $P_s(t, \omega)$ . Then the linearly frequency modulated signal  $s(t)e^{j\beta t^2/2}$  will have covariance  $R_s(t, \tau)e^{j\beta t\tau}$  and corresponding nonnegative TF spectrum  $P_s(t, \omega - \beta t)$ . More generally, let  $x(t) = s(t)e^{j\phi(t)}$ , where  $s(t)$  is locally stationary with symmetric covariance function  $R_s(t, \tau)$  from equation (16). Then the covariance of  $x(t)$  is

$$R_x(t, \tau) = g(t)f(\tau)e^{j(\phi(t+\tau/2)-\phi(t-\tau/2))}. \quad (18)$$

By making use of the principle of stationary phase [14], it can be shown [13] that the TF spectrum of  $x(t)$  is given by:

$$P_x(t, \omega) = g(t)F(\omega - \phi'(t)) = P_s(t, \omega - \phi'(t)). \quad (19)$$

Thus, a frequency modulated locally stationary (FMLS) process will have a TF spectrum equal to that of the locally stationary process centered around the instantaneous frequency of the FM. The generalization can be taken one step further to define a *composite FMLS process*, consisting of a sum of statistically independent FMLS processes. The composite signal will also have a nonnegative TF spectrum equal to the sum of the spectra of the individual processes.

However, when  $s(t)$  is an FMLS process,  $P(t, \omega)$  will most certainly *not* be constant over  $\Delta T \times \Delta W$ , and equation (12) will fail to be valid. In this case, the smoothing region  $\Delta T \times \Delta W$  must be oriented to match the FM of the signal. This reorientation is equivalent to matching the spectrogram window to the FM of the signal. This matching can be accomplished by using a frequency modulated window in the original STFT computation. However, in signals with multiple FM rates, as in a composite FMLS signal, this adaptation must be performed locally in TF, as discussed next.

### 4. LOCALLY MATCHED MULTITAPER SPECTROGRAMS

To locally demodulate the spectrograms, it is first necessary to construct a reliable estimate of the local FM, which is denoted by  $\beta(t, \omega)$ . Letting the TF dependence be implicit,  $\beta$  can be estimated by computing a local covariance of the multitaper spectrogram normalized by the time spread:  $\langle (t - \bar{t})(\omega - \bar{\omega}) \rangle / \langle (t - \bar{t})^2 \rangle$ , where  $\bar{t}$  and  $\bar{\omega}$  are the local average time and frequency, respectively; their dependence on  $t$  and  $\omega$  is implied. The covariance is computed by integrating over a finite region of the TF plane  $\Delta T \times \Delta W$  as a two-dimensional

sliding window to provide an estimate of  $\beta$  as a function of  $t$  and  $\omega$ :

$$\beta(t, \omega) = \frac{\int_{\Delta T} \int_{\Delta W} (t - \hat{t} - \bar{t})(\omega - \hat{\omega} - \bar{\omega}) P(\hat{t}, \hat{\omega}) d\hat{t} d\hat{\omega}}{\int_{\Delta T} \int_{\Delta W} (t - \hat{t} - \bar{t})^2 P(\hat{t}, \hat{\omega}) d\hat{t} d\hat{\omega}}; \quad (20)$$

$\bar{t}$  and  $\bar{\omega}$  are computed similarly. Integrating over a larger region will provide better variance properties at the expense of possible bias due to multiple signal components with differing FM rates lying within the area of integration.

Once  $\beta(t, \omega)$  has been estimated, each STFT  $S_k(t, \omega)$  is dechirped by locally convolving it with the Fourier transform of  $e^{j\beta(t, \omega)\tau^2/2}$ :

$$S_k^\beta(t, \omega) = \int S_k(t, \omega - \theta) e^{-j\theta^2/2\beta(t, \omega)} d\theta. \quad (21)$$

This convolution is shift-variant; at each frequency, a new  $\beta$  must be used. This convolution is equivalent to matching the STFT to the local chirp rate. While this convolution at first would appear to be an  $O(N^2)$  operation, it can actually be implemented much more efficiently. The equivalent chirp in the time domain is of length  $T$ , the length of the STFT window. The Fourier transform of this finite-length chirp will then have bandwidth  $\beta T$ . Thus, if the average bandwidth of the various FM components is  $M = \beta T$  bins, an STFT with  $N$  frequency samples can be dechirped with only  $NM$  multiplies per time slice, comparable to the computational complexity of the STFT itself. Once all of the  $S_k(t, \omega)$  are dechirped, the multitaper estimate is constructed as usual.

## 5. F-TEST FOR FREQUENCY-MODULATED TONES

The validity of the multitaper estimate rests on the assumption that the TF spectrum is smoothly varying over  $\Delta T \times \Delta W$ . This assumption is violated when spectral lines (FM or otherwise) are present in the signal. In this case, it is necessary to estimate the tones and remove them from the signal. Ordinarily, estimating a tone with unknown FM would be extremely difficult. This task is made easier, however, by the local matching described above. Once the individual STFT's  $S_k(t, \omega)$  have been adapted to local FM, any frequency modulated tones in the signal will behave exactly as a stationary tone would behave in a non-adapted STFT. As a result, an F-test for the existence of any FM tones in the TF spectrum can be defined by directly extending Thomson's approach in the stationary case. The expected value of the  $k^{th}$  dechirped STFT for an FM tone  $\mu e^{j\phi(t)}$  with instantaneous frequency  $\omega = \phi'(t)$  is:

$$E[S_k(t, \omega)] = \mu U_k(0). \quad (22)$$

The mean can then be estimated via regression:

$$\hat{\mu}(t, \omega) = \frac{\sum_{k=1}^K U_k(0) S_k(t, \omega)}{\sum_{k=1}^K U_k^2(0)}. \quad (23)$$

The variance of this estimate is equal to the background TF spectrum minus the spectral line, which is:

$$P(t, \omega) = \frac{1}{K-1} \sum_{k=1}^K |S_k(t, \omega) - \hat{\mu}(t, \omega) U_k(0)|^2. \quad (24)$$

The F-test at time  $t$  is then given by the ratio of the power of the spectral line and that of the background spectrum:

$$F(t, \omega) = \frac{(K-1) |\hat{\mu}(t, \omega)|^2 \sum_{k=1}^K U_k^2(0)}{\sum_{k=1}^K |S_k(t, \omega) - \hat{\mu}(t, \omega) U_k(0)|^2}. \quad (25)$$

Under the null hypothesis, the test quantity at a single time is the ratio of two  $\chi^2$  random variables with 2 and  $2(K-1)$  degrees of freedom. For a signal of length  $T$  and an STFT of order  $N$ , there will be  $T/N$  independent blocks of data. Thus, the final F-test will be a ratio of  $\chi^2$  random variables with  $2T/N$  and  $2(K-1)T/N$  degrees of freedom, integrated along the contour specified by  $\omega = \phi'(t)$ :

$$F(\phi'(t)) = \frac{(K-1) \sum_{t=1}^T |\hat{\mu}(t, \phi'(t))|^2 \sum_{k=1}^K U_k^2(0)}{\sum_{t=1}^T \sum_{k=1}^K |S_k(t, \phi'(t)) - \hat{\mu}(t, \phi'(t)) U_k(0)|^2}. \quad (26)$$

If the F-test achieves the specified confidence level, the tone should be removed by subtracting from the STFT's prior to forming the TF spectrum, then added into the representation as an impulse:

$$P(t, \omega) = \hat{\mu}(t, \omega) \delta(\omega - \phi'(t)) + \frac{1}{K} \sum_{k=1}^K |S_k(t, \omega) - \hat{\mu}(t, \omega) U_k(\omega - \phi'(t))|^2. \quad (27)$$

Matching the STFTs to the local FM greatly simplifies the F-test. With no matching, the STFT of an FM tone will be spread according to the sweep rate, and will thus have a functional form dependent on  $\beta$ . After matching, the FM tone will have the same response as a stationary tone in an unmatched STFT. Thus, the expression for  $\mu$  in equation (23) can be used for all FM rates. The procedure for testing for an FM tone is then a four-step process: compute the test statistic  $F(t, \omega)$  over time and frequency; find candidate contours  $\omega(t) = \phi'(t)$  in  $F(t, \omega)$ ; compute  $F(\phi'(t))$ ; and test its significance.

## REFERENCES

- [1] D. J. Thomson, "Spectrum estimation and harmonic analysis," *Proceedings of the IEEE*, vol. 70, pp. 1055–1096, 1982.
- [2] D. Thomson and A. Chave, "Jackknifed error estimates for spectra, coherences, and transfer functions," in *Advances in Spectrum Analysis and Array Processing, Vol. I* (S. Haykin, ed.), pp. 58–113, Prentice-Hall, 1991.
- [3] M. Bayram and R. G. Baraniuk, "Multiple window time-frequency analysis," in *Proceedings of the IEEE-SP International Symposium on Time-Frequency and Time-Scale Analysis*, (Paris, France), pp. 173–176, June 1996.
- [4] I. Daubechies, "Time-frequency localization operators: a geometric phase space approach," *IEEE Transactions on Information Theory*, vol. 34, no. 4, pp. 605–612, 1992.
- [5] F. Çakrak and P. Loughlin, "Multiple window non-linear time-varying spectral analysis," *IEEE Transactions on Signal Processing*, 1999. submitted.
- [6] J. Pitton, "Nonstationary spectrum estimation and time-frequency concentration," in *IEEE Conference on Acoustics, Speech, and Signal Processing*, vol. IV, pp. 2425–2428, IEEE, 1998.
- [7] J. Pitton, "Time-frequency spectrum estimation: an adaptive multitaper method," in *IEEE Int. Sym. Time-Frequency and Time-Scale Analysis*, (Pittsburgh, PA), pp. 665–668, 1998.
- [8] J. Pitton, "Adaptive multitaper time-frequency spectrum estimation," in *SPIE Advanced Sig. Proc. Algs. Archs., Impl. VII*, 1999.
- [9] M. Priestley, *Spectral Analysis of Time Series*. Academic Press - London, 1981.
- [10] J. Pitton, "Linear and quadratic methods for positive time-frequency distributions," in *IEEE Conference on Acoustics, Speech, and Signal Processing*, vol. 5, pp. 3649–3652, IEEE, 1997.
- [11] P. Flandrin, "On the positivity of the wigner-Ville spectrum," *Signal Processing*, vol. 11, no. 2, 1985.
- [12] R. Silverman, "Locally stationary random processes," *IRE Transactions on Information Theory*, vol. IT-3, September 1957.
- [13] J. Pitton, "The statistics of time-frequency analysis," *Journal of the Franklin Institute*, 2000. To appear.
- [14] E. Key, E. Fowle, and R. Haggarty, "A method of designing signals of large time-bandwidth product," *IRE Intern. Conv. Record*, vol. IV, 1961.

# OPTIMAL SUBSPACE SELECTION FOR NON-LINEAR PARAMETER ESTIMATION APPLIED TO REFRACTIVITY FROM CLUTTER

Shawn Kraut and Jeffrey Krolik

Department of Electrical and Computer Engineering  
Duke University, Box 90291  
Durham, NC 27708-0291

E-mail: kraut@ee.duke.edu, jk@ee.duke.edu

## ABSTRACT

We consider the problem of constructing an optimal reduced-rank subspace for parameter estimation, in models where the data is a non-linear function of the parameters. The solution which minimizes mean-squared error is a compromise between the prior distribution, and the measurement model, reducing to the Karhunen-Loeve Transform when only the prior is considered. The measurement model determines which parameters the measured data is less sensitive to, and which are therefore less estimatable. Our approach obtains parameterizations in which the influence of these parameters is reduced, so that limited resources may be allocated to more estimatable features. We apply it to the problem of estimating index-of-refraction profiles from sea-surface clutter data.

## 1. INTRODUCTION

In this paper we will consider the problem of constructing a reduced-dimension subspace in which to search for parameter estimates  $\hat{\theta}$ . Non-linear models of the form  $y = L(\theta) + n$  are considered, where the measured data  $y$  depends on the parameter set  $\theta$  through the non-linear model  $L(\cdot)$ , and is corrupted by additive noise  $n$ . We will discuss this problem in the specific case of estimating the tropospheric index of refraction profile from clutter returns received from ship-based microwave radars [1]. In the "refractivity from clutter" (RFC) problem, the data  $y$  consists of clutter returns across range, and the description of propagation through the refractivity profile yields the non-linear model.

We ask the following question: what is the optimal *reduced-rank* basis for searching for estimates of

the parameter set? From an engineering standpoint, estimating the full refractivity profile would require a search through a large-dimensional parameter space, and would be too computationally slow for real-time estimation of a dynamically varying profile. From a modeling standpoint, we are interested in what reduced parameterizations one should be estimating.

The Karhunen-Loeve Transform (KLT) describes the optimal reduced-rank linear subspace for minimizing compression or representation error [2], by considering the prior statistical distribution of the parameters. The subspace is constructed from the dominant eigenvectors of the prior covariance matrix of the parameter vector,  $R_{\theta\theta} = E[\theta\theta^T]$  (with the mean of  $\theta$  subtracted out). The limitation of the KLT is that it does not incorporate the *estimation* problem: what parameterizations can be estimated from the data with the smallest estimation error? If one were to consider estimation error alone, then one would build the reduced-rank search space from the *model*  $L(\cdot)$ , ignoring the prior. But the resulting parameter basis functions might not represent well the natural distribution of the parameters. In the RFC example, profiles that are built from such a basis will not necessarily look like natural, typically observed index-of-refraction profiles (for an example, see Figure 1).

The optimal basis, in a MSE sense, is a compromise between the two considerations of estimation and representation error. What is this basis? In the case of *linear* models ( $y=L\theta+n$ ), the problem has been investigated and solved in two contexts. Examining Wiener filters, in the form  $R_{\theta y}R_{yy}^{-1}$ , Scharf found that the optimal (minimum mean-square error) *reduced-rank Wiener filter* is given by truncating the singular-value decomposition (SVD) of  $R_{\theta y}R_{yy}^{-\frac{1}{2}}$ , to give  $\text{trunc}[(R_{\theta y}R_{yy}^{-\frac{1}{2}})R_{yy}^{-\frac{1}{2}}]$  (see [3], p.330, and [4]). More recently, Hua, et. al., suggested the *generalized KLT (GKLT)*, constructed from the dominant eigenvectors of  $R_{\theta y}R_{yy}^{-1}R_{y\theta}$  (see [5,

This work was supported by SPAWAR Systems Center, San Diego, under contract No. N66001-97-D-5028. Presented at the 10th IEEE Workshop on Statistical Signal and Array Processing, Pocono Manor, Pennsylvania, August 14-16, 2000.

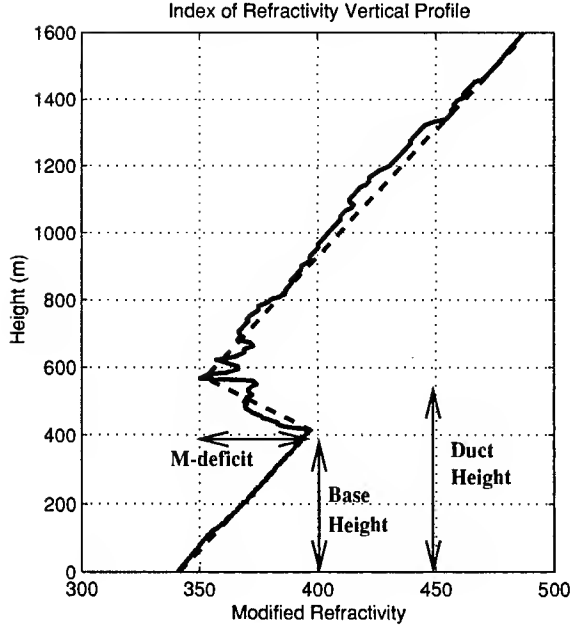


Figure 1: A typical tropospheric **index of refraction** profile, with a tri-linear shape characterized in part by base height, duct height, and M-deficit.

6]).

We are considering the same problem, but in the context of *non-linear* models. Furthermore, in the RFC case we will discuss, a closed-form analytical model for  $\underline{L}(\underline{\theta})$  is not available; the clutter return  $\underline{y}$  that would result from a given profile  $\underline{\theta}$  must be computed numerically. How then do we find the optimal reduced-rank parameter basis?

## 2. ORTHOGONALITY CONDITIONS

Generally, one seeks a solution  $\hat{\underline{\theta}}$  that maximizes some objective function  $\mathcal{L}$ :

$$\max_{\hat{\underline{\theta}}} \mathcal{L}(\underline{y}, \underline{L}(\hat{\underline{\theta}})) \rightarrow \hat{\underline{\theta}}. \quad (1)$$

For example, for a MAP (maximum *a posteriori*) estimator,  $\mathcal{L}$  maximizes the posterior probability density function for  $\underline{y}$ . In this work we are seeking to identify linear, reduced-rank parameterizations in the form  $\hat{\underline{\theta}}_r = \underline{U}_r \underline{b}$ . Here  $\underline{U}_r$  is a “tall” matrix with orthonormal columns, i.e.  $\underline{U}_r^\dagger \underline{U}_r = \underline{I}$ . The problem is then reduced to searching over candidate values of  $\underline{b}$ :

$$\max_{\underline{b}} \mathcal{L}(\underline{y}, \underline{L}(\underline{U}_r \underline{b})) \rightarrow \hat{\underline{\theta}}_r = \underline{U}_r \hat{\underline{b}}, \quad (2)$$

and the basic question is, how do we choose  $\underline{U}_r$  ?

Some useful results can be obtained by assuming the following two axioms: both the full-rank and reduced-rank estimators are uncorrelated with (orthogonal to) the error of the full-rank estimator:

$$(A) \ E[(\underline{\theta} - \hat{\underline{\theta}})\hat{\underline{\theta}}^\dagger] = \underline{0}; \text{ and } (B) \ E[(\underline{\theta} - \hat{\underline{\theta}})\hat{\underline{\theta}}_r^\dagger] = \underline{0}.$$

(3)

The first condition is strictly true for the conditional mean (CM) estimator, which is also the Minimum Mean-Squared Error (MMSE) estimator, and for the Linear MMSE estimator (Wiener filter). It can be shown that the second condition is strictly true if  $\hat{\underline{\theta}}$  is constructed from the the MMSE estimator or LMMSE estimator, and  $\hat{\underline{\theta}}_r$  is constructed from the same type of estimator of  $\underline{\beta} = \underline{U}_r^\dagger \underline{\theta}$ . This condition basically excludes  $\hat{\underline{\theta}}_r$  from bringing in side information about  $\underline{\theta}$  that is not present in  $\hat{\underline{\theta}}$ . (In simple terms, we don’t have the situation where  $\hat{\underline{\theta}}$  is poor estimator, while  $\hat{\underline{\theta}}_r$  is simultaneously based on a good estimator.)

A consequence of this condition is that the error correlation of  $\hat{\underline{\theta}}_r$  is greater than that of  $\hat{\underline{\theta}}$ :

$$\underline{Q}_r = \underline{Q} + E[(\hat{\underline{\theta}} - \hat{\underline{\theta}}_r)(\hat{\underline{\theta}} - \hat{\underline{\theta}}_r)^\dagger] \geq \underline{Q}, \quad (4)$$

where  $\underline{Q} = E[(\underline{\theta} - \hat{\underline{\theta}})(\underline{\theta} - \hat{\underline{\theta}})^\dagger]$ . If we seek the reduced-rank estimator that minimizes the residual MSE (trace of  $E[(\hat{\underline{\theta}} - \hat{\underline{\theta}}_r)(\hat{\underline{\theta}} - \hat{\underline{\theta}}_r)^\dagger]$ ), it can be shown that the error correlation can be rewritten as

$$\underline{Q}_r = \underline{Q} + (\underline{I} - \underline{P}_r)\underline{R}_{\hat{\underline{\theta}}\hat{\underline{\theta}}}(\underline{I} - \underline{P}_r), \quad (5)$$

where  $\underline{R}_{\hat{\underline{\theta}}\hat{\underline{\theta}}} = E[\hat{\underline{\theta}}\hat{\underline{\theta}}^\dagger]$  is the *estimator* correlation, and  $\underline{P}_r = \underline{U}_r \underline{U}_r^\dagger$  is the projection onto the reduced-rank subspace. Using the same argument as that taken for the KLT, the reduced-rank subspace is then constructed from the dominant eigenvectors of  $\underline{R}_{\hat{\underline{\theta}}\hat{\underline{\theta}}}$ . It should be noted that in the case of the linear model, this result reduces to the “Generalized KLT” discussed in [3, 4, 5, 6]; i.e.  $\underline{R}_{\hat{\underline{\theta}}\hat{\underline{\theta}}}$  becomes  $\underline{R}_{\underline{y}\underline{\theta}}\underline{R}_{\underline{\theta}\underline{\theta}}^{-1}\underline{R}_{\underline{\theta}\underline{y}}$ .

This solution is intuitively pleasing: to find a reduced-rank subspace to search for parameter estimates, search the subspace where the full-rank estimates naturally tend to lie. Also, note that a consequence of the first orthogonality condition is that the *a priori* covariance of the parameter vector  $\underline{\theta} = (\underline{\theta} - \hat{\underline{\theta}}) + \hat{\underline{\theta}}$  can be decomposed into the correlation of the error  $(\underline{\theta} - \hat{\underline{\theta}})$ , and the correlation of the full-rank estimator  $\hat{\underline{\theta}}$ . This observation can be written in the form of a “Pythagorean Theorem”:

$$\begin{aligned} \underline{R}_{\underline{\theta}\underline{\theta}} &= \underline{Q} + \underline{R}_{\hat{\underline{\theta}}\hat{\underline{\theta}}} \\ \rightarrow \underline{R}_{\hat{\underline{\theta}}\hat{\underline{\theta}}} &= \underline{R}_{\underline{\theta}\underline{\theta}} - \underline{Q}. \end{aligned} \quad (6)$$

In this formulation, we should take the dominant eigenvectors of the difference between the *a priori* covariance and the full-rank error correlation, which reduces to the KLT in the limit that the error correlation becomes small.

### 3. CONSTRUCTING THE SUBSPACE IN PRACTICE

Estimating the covariance matrix  $\mathbf{R}_{\hat{\theta}\hat{\theta}}$  over the full parameter space may be computationally intensive in practice, limited by the computation time of the propagation model  $\underline{L}(\cdot)$ , over a set of values of  $\underline{\theta}$  (either grid points or realizations). Recently, closed-form expressions were obtained for the Fisher information matrix in the RFC estimation problem [7], which could in principle be used to approximate the full-rank error correlation in Equation 6, i.e.  $\mathbf{Q} \gtrsim \mathbf{J}^{-1}$ . However, this approach is infeasible if the dimension of the initial parameter vector  $\underline{\theta}$  is too high, since the multi-dimensional numerical differentiation for the estimate of  $\mathbf{J}$  requires the evaluation of the nonlinear function  $\underline{L}(\cdot)$  on a number of grid points that increases quadratically with the dimension.

An alternate approach taken here is to obtain samples for a sample covariance matrix estimate of  $\mathbf{R}_{\hat{\theta}\hat{\theta}}$ , where each sample is an approximate conditional-mean estimate  $\hat{\underline{\theta}}$ , formulated as follows:

$$\begin{aligned} \hat{\underline{\theta}}_{CM}(\underline{y}) &= \int d\underline{\theta} \underline{\theta} f(\underline{\theta}|\underline{y}) = \frac{\int d\underline{\theta} \underline{\theta} f(\underline{\theta}, \underline{y})}{\int d\underline{\theta} f(\underline{\theta}, \underline{y})} \\ &= \frac{\int d\underline{\theta} \underline{\theta} f(\underline{\theta}|\underline{y}) f(\underline{\theta})}{\int d\underline{\theta} f(\underline{\theta}|\underline{y}) f(\underline{\theta})} \approx \frac{\sum_i \underline{\theta}_i f(\underline{\theta}_i|\underline{y})}{\sum_i f(\underline{\theta}_i|\underline{y})} \\ &= \sum_i \underline{\theta}_i w(\underline{y}, \underline{\theta}_i), \quad w(\underline{y}, \underline{\theta}_i) = \frac{f(\underline{y}|\underline{\theta}_i)}{\sum_i f(\underline{y}|\underline{\theta}_i)} \quad (7) \end{aligned}$$

where  $\underline{\theta}_i$  are samples drawn from the prior  $f(\underline{\theta})$ , i.e. historical data, and  $w(\underline{y}, \underline{\theta}_i)$  is a normalized weighting factor, proportional to the likelihood. So an estimate is obtained by averaging over historical profiles that are weighted by their likelihood of producing the data  $\underline{y}$ .

### 4. APPLICATION: ESTIMATION OF TROPOSPHERIC REFRACTIVITY PROFILES

To evaluate this approach to rank-reduction, we used profiles from the VOCAR data set, taken at three sites off the coast of southern California in 1993 [1]. The most straightforward way to apply the KLT approach is to simply take the profiles of M-values (modified refractivity) over a uniform height grid, and concatenate them into the columns of a data matrix  $\Theta$ , and

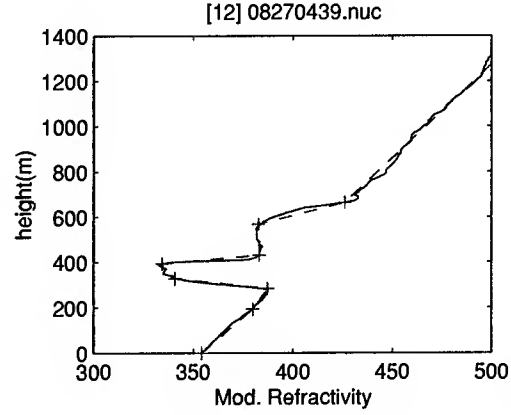


Figure 2: An “octo-linear” fit of a profile, consisting of eight linear segments.

use the resulting sample covariance of  $\mathbf{R}_{EOF}$  to generate dominant eigenvectors/EOFs (extended orthogonal functions), and in order to generate new random profiles for analysis. Unfortunately, a Gaussian random model with covariance  $\mathbf{R}_{EOF}$  fails to reproduce the characteristic tri-linear shape of observed profiles (Figure 1, the second linear segment is responsible for the downward refraction that causes ducting behavior). In particular, the height of the duct (height of the first two segments in the tri-linear profile) may vary considerably, and averaging over an ensemble of observed profiles tends to suppress the key feature of the duct; it is “washed-out” in the sample mean (not shown here). In addition, profiles synthesized from the sample mean and  $\mathbf{R}_{EOF}$  tend to have many mini-ducts over the entire height range, features not observed in real data.

To formulate a random model that synthesizes realistic profiles, and at the same time formulate an initial profile parameterization, we fit each historical profile to a profile consisting of eight linear segments (i.e., an “octo-linear” profile), as shown in Figure 2. This procedure fits the profile to a length-17 parameter vector  $\underline{\theta}$ , corresponding to the heights of the eight segments, the widths of the eight segments (or M-deficits), and the the M-value at zero height (sea level).

The key characteristic of this fit is that it is *feature-based*: referring to Figures 1 and 2, the top of the second segment was generally chosen to correspond to the middle of the duct (the first two segments accounting for base-height), and the top of the fourth segment was chosen to correspond to the top of the duct (the first four segments accounting for duct height). (For the results shown here, the fit was obtained manually.) To reduce a spurious source of variance in these parameters, the historical profiles were edited to remove those

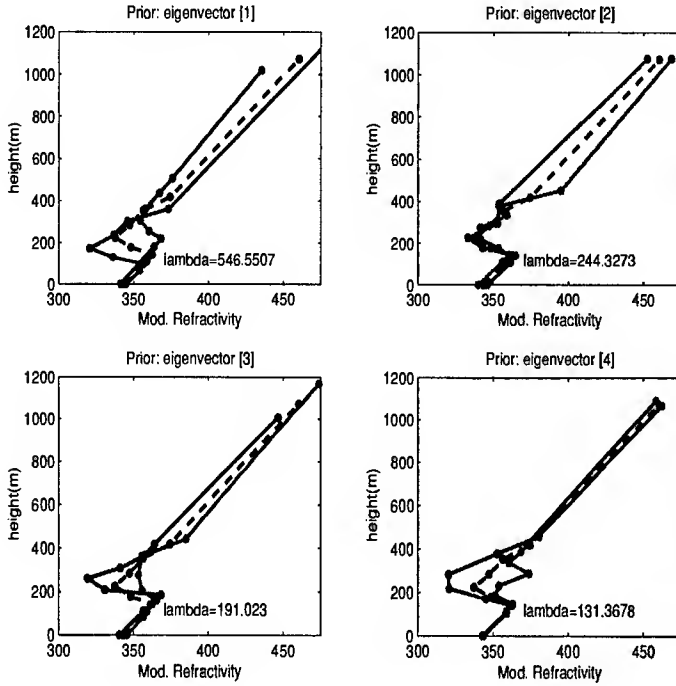


Figure 3: The first four dominant eigenvectors of the prior covariance,  $\mathbf{R}_{\theta\theta}$ . Each panel contains three plots, the profiles corresponding to (1) the mean (dashed), (2) and (3) the mean  $\pm$  the eigenvector (scaled by the same constant in all four panels).

for which the main duct feature was not identifiable (such as profiles that looked basically linear, with no apparent duct).

As might be expected, profiles synthesized from a multivariate Gaussian model on feature-based *parameters*, rather than on the raw profiles, are more realistic in terms of reproducing the gross shape of a typical profile, including the main duct. To insure positivity, a log-normal model was used on the heights (the appropriateness of which was verified by inspection of histograms from real data). The multivariate-normal model was then applied to the vector of  $\log(\text{heights})$  and M-deficits.

Interestingly, the resulting mean, the dashed line in the panels of Figure 3, looks very tri-linear. The influence of the dominant eigenvectors of the prior covariance  $\mathbf{R}_{\theta\theta}$  are depicted by the solid lines in Figure 3. The first eigenvector corresponds to increasing base-height while decreasing M-deficit in the tri-linear model. The second has a lot of energy going into shrinking and expanding the length of the top segment. This by itself is a strict degeneracy: scaling of the length of the final segment has no effect on the profile and no effect on the clutter measurements used to

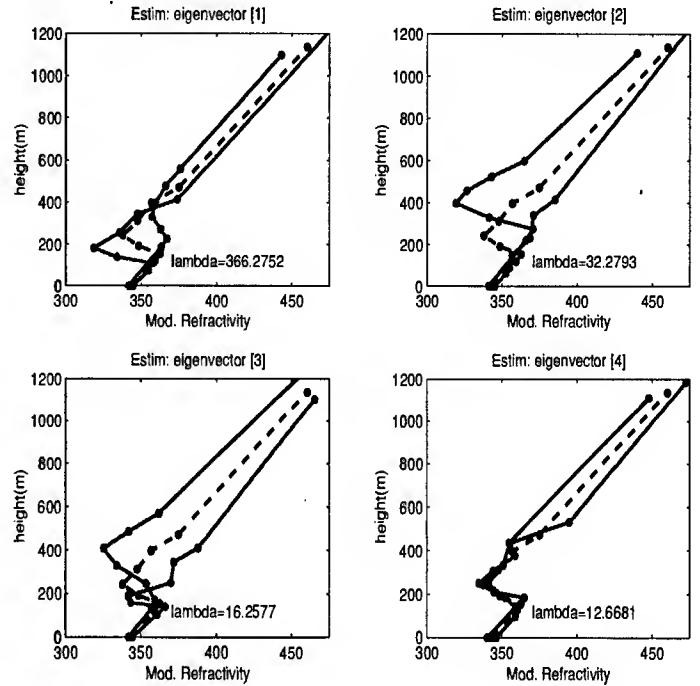


Figure 4: The first four dominant eigenvectors of the sample estimator covariance,  $\mathbf{R}_{\hat{\theta}\hat{\theta}}$ .

estimate the profile; it is only an artifact of the initial parameterization scheme. Furthermore, for the measurement method presumed for this study, measurement of sea-surface clutter strength across range, variations in the top-half of the the profile constitute an effective estimation degeneracy, since they have little effect on the ducting behavior and measured surface clutter, and are therefore difficult to estimate.

We used a sample covariance approach to approximate the estimator covariance  $\mathbf{R}_{\hat{\theta}\hat{\theta}}$  of Equation 5. A likelihood function  $f(\mathbf{y}|\underline{\theta})$  is easily obtainable, as a function of the propagation loss  $\underline{L}(\underline{\theta})$  from the transmitter to the sea surface, across range (where the dimension of  $\mathbf{y}$  and  $\underline{L}$  is the number of range cells). The problem is that the PE (parabolic equation) numerical propagation of the field is time intensive, severely limiting the number of parameter values  $\underline{\theta}_i$  at which the propagation loss can be evaluated.

To generate samples for a sample covariance, we computed approximate conditional-mean based estimates  $\hat{\underline{\theta}}$ , based on the weighted sum of Equation 7. In practice, direct implementation of Equation 7 failed, since the number of samples  $\underline{\theta}_i$  (10,000) was too small to adequately sample the likelihood function, forcing one weight  $w_i$  to be unity, and the rest to be zero. This effect was ameliorated by increasing the standard deviation of the likelihood function by a factor of 35.



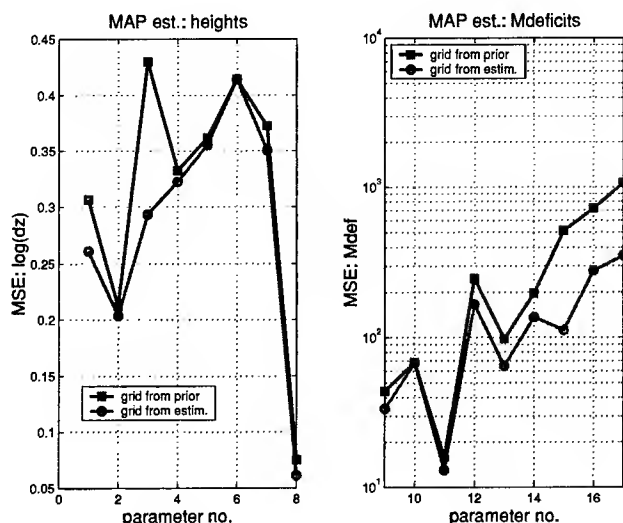


Figure 5: The mean-square-error (MSE) for the MAP estimator over a grid (1) based on the prior covariance (KLT) and (2) based on the estimator covariance.

The weighted sum can be interpreted as summing over different profiles that reproduce well the observed data. This in turn has the effect of averaging over, or "washing out", variations corresponding to the degeneracies discussed above, which have less impact on the measurement, and which are therefore less estimatable.

The sample covariance of the resulting estimates has the eigenvectors shown in Figure 4. These eigenvectors are qualitatively preferable those of the prior covariance, in terms of their physical interpretations: the first corresponds to increasing duct height while decreasing M-deficit, and the second to increasing base height with increasing M-deficit. Note that the second eigenvector of the prior covariance, in Figure 3 (with energy going into scaling the top segment), is here most closely approximated by the fourth eigenvector. So the energy going into this degenerate, non-estimatable feature has been reduced.

To quantitatively compare this parameterization with that of the KLT, two grids consisting of 6000 points were constructed from the dominant three eigenvectors of the prior and estimator covariance, respectively. The number of grid points was determined by the relative energy of the eigenvectors, as reflected by the eigenvalues;  $25 \times 16 \times 15$  and  $40 \times 15 \times 10$  grids were chosen for the prior and estimator covariance eigenvectors, respectively. The mean-square-error decreases when MAP estimates are found over the grid based on the estimator covariance; see Figure 5.

## 5. CONCLUSIONS

In this paper we have discussed the problem of describing the lower-dimensional parameterization of an unknown parameter set that is optimal in the sense of minimizing mean-squared error. This description, in terms of a reduced-rank subspace, depends on both the measurement model by which the data depends on the parameters and on the *a priori* distribution of the parameters. It can be viewed as a generalization of the Karhunen-Loeve Transform, which considers only the prior. The initial parameterization and the nature of the measurement model may contain parameters which are degenerate in the sense that they have less impact on the measured data. The aim of the approach presented in this paper is to seek parameterizations in which the strength of these parameters is decreased, so that the reduced-dimension parameterization emphasizes more estimatable features. We have evaluated this procedure for the application of estimating index of refraction profiles from clutter returns, where it produces more physically meaningful reduced-rank basis functions, and decreases mean-squared-error relative to the KLT basis.

## REFERENCES

- [1] T. Rogers, "Effects of the variability of atmospheric refractivity on propagation estimates," *IEEE Trans. Antennas Propagat.*, vol. 44, pp. 460-465, April 1996.
- [2] C. W. Therrien, *Discrete Random Signals and Statistical Signal Processing*, Signal Processing Series, Ed. A. V. Oppenheim. Prentice Hall, 1992.
- [3] L. L. Scharf, *Statistical Signal Processing*, Addison-Wesley, 1991.
- [4] L. L. Scharf, "The SVD and reduced rank signal processing," *Signal Processing*, vol. 25, pp. 113-133, 1991.
- [5] Y. Hua and W. Liu, "Generalized Karhunen-Loeve transform," *IEEE Signal Processing Letters*, vol. 5, no. 6, pp. 141-142, June 1998.
- [6] Y. Hua and M. Nikpour, "Computing the reduced rank Wiener filter by IQMD," *IEEE Signal Processing Letters*, submitted.
- [7] J. Tabrikian, "Theoretical performance limits on tropospheric refractivity estimation using point-to-point microwave measurements," *IEEE Trans. Antennas Propagat.*, vol. 47, no. 11, pp. 1727-1734, November 1999.

# MAP MODEL ORDER SELECTION RULE FOR 2-D SINUSOIDS IN WHITE NOISE

Mark A. Kliger and Joseph M. Francos

Department of Electrical and Computer Engineering  
Ben-Gurion University  
Beer-Sheva 84105, Israel.

## ABSTRACT

We consider the problem of jointly estimating the number as well as the parameters of two-dimensional sinusoidal signals, observed in the presence of an additive white Gaussian noise field. Existing solutions to this problem are based on model order selection rules, derived for the parallel one-dimensional problem. These criteria are then adapted to the two-dimensional problem using heuristic arguments. Employing asymptotic considerations, we derive in this paper a maximum a-posteriori (MAP) model order selection criterion for jointly estimating the parameters of the two-dimensional sinusoids and their number.

## 1. INTRODUCTION

From the 2-D Wold-like decomposition we have that any 2-D regular and homogeneous discrete random field can be represented as a sum of two mutually orthogonal components: a purely-indeterministic field and a deterministic one. The purely-indeterministic component has a unique white innovations driven moving average representation. The deterministic component is further orthogonally decomposed into a harmonic field and a countable number of mutually orthogonal evanescent fields. In this paper we consider a special case of the foregoing general problem. More specifically, we consider the problem of jointly estimating the number as well as the parameters of the sinusoidal signals comprising the harmonic component of the field, in the presence of the purely-indeterministic component, assumed here to be a white noise field.

A solution to this problem is an essential component in many image processing and multimedia data processing applications. For example, in indexing and

retrieval systems of multimedia data that employ the textural information in the imagery components of the data, *e.g.*, [7], the identification of similar textured surfaces as being such, is highly sensitive to errors in estimating the orders of the models of the deterministic components of the textures. More specifically, in this approach the 2-D Wold decomposition based parametric model of each textured segment of the image also serves as the index of this segment. Therefore an accurate and robust procedure for estimating the orders as well as the parameters of the models of the deterministic components of the textures is an essential component in any such indexing and retrieval system. Similar requirements are posed by parametric content-based image coding and representation methods.

The same type of problem, *i.e.*, joint estimation of the model order and parameters for a sum of 2-D sinusoidal signals observed in additive noise, naturally arises in processing 2-D SAR data. In this problem however the observed random field is complex valued, where for each scatterer one frequency parameter corresponds to the range information, while the second frequency parameter is the Doppler. The complex valued amplitude of each such exponential is proportional to the radar cross section of the target.

Many algorithms have been devised to estimate the parameters of sinusoids observed in additive white Gaussian noise. Most of the algorithms assume that the number of sinusoids is *a-priori* known. However this assumption does not always hold in practice. Hence, in the past two decades the model order selection problem has received considerable attention. In general, model order selection rules are based (directly or indirectly) on three popular criteria: Akaike information criterion (AIC), the minimum description length (MDL) and the maximum a-posteriori probability (MAP) criterion. All these criteria have a common form in that they comprise two terms: a data term and a penalty term, where the data term is the log-likelihood function evaluated for the assumed model. However, most of the papers

This work was supported in part by the Israel Ministry of Science under Grant 1233198.

dedicated to this problem discuss the model order selection problem for various models of one-dimensional signals, while the problem of modeling multidimensional fields has received considerably less attention. Djuric, [1], proposed a MAP order selection rule for 1-D sinusoids observed in additive white noise. Kavalieris and Hannan, [4], prove the strong consistency of a criterion, that indirectly employs the MDL principle. In this framework the observation noise is modeled as an autoregression of an unknown order. In the special case where the noise process in [4] is assumed to be a white noise process, the resulting criterion is identical to the MAP criterion derived in [1]. Stoica *et al.*, [5] proposed the cross-validation selection rule and demonstrated its asymptotic equivalence to the Generalized Akaike Information Criterion (GAIC). In [6] this criterion is applied to the 2-D problem as well, where the penalty term is proportional to the total number of unknown parameters, exactly as in the 1-D case. In this paper we derive a MAP model order selection criterion for jointly estimating the number and the parameters of two-dimensional sinusoids observed in additive white noise.

The paper is organized as follows. In Section 2 we define our notations, while in Section 3 we formally define the MAP model order selection problem. The MAP model order selection criterion is derived in Section 4. Finally, in Section 5 we provide some numerical examples and Monte-Carlo simulations to better illustrate the performance of the proposed criterion.

## 2. NOTATIONS AND DEFINITIONS

The considered random field is composed of an harmonic field embedded in Gaussian noise. Let  $\{y(n, m)\}$  where  $(n, m) \in U$  and  $U = \{(n, m) \mid 0 \leq n \leq S-1, 0 \leq m \leq T-1\}$ , be the observed  $S \times T$  real valued data field. The elements of  $y(n, m)$  may be represented as

$$y(n, m) = h(n, m) + u(n, m). \quad (1)$$

The field  $\{u(n, m)\}$  is the 2-D zero mean, Gaussian white noise field with variance  $\sigma^2$ . The field  $\{h(n, m)\}$  is the harmonic random field

$$h(n, m) = \sum_{i=1}^k C_i \cos(n\omega_i + m\nu_i) + G_i \sin(n\omega_i + m\nu_i) \quad (2)$$

where  $k$  denotes the number of sinusoidal components in the data model, and  $(\omega_i, \nu_i)$  is the spatial frequency of the  $i$ th component. The  $C_i$ 's and  $G_i$ 's are the amplitudes of the sinusoidal components in the observed realization.

Let us define the following matrix notations:

$$\mathbf{y} = [y(0, 0), \dots, y(0, T-1), y(1, 0), \dots, \dots, y(S-1, T-1)]^T \quad (3)$$

The vectors  $\mathbf{u}$  and  $\mathbf{h}$  are similarly defined. Rewriting (1) we have  $\mathbf{y} = \mathbf{h} + \mathbf{u}$ . Let  $\mathbf{\Lambda}$  denote the covariance matrix of  $\mathbf{y}$ . Thus

$$\mathbf{\Lambda} = \sigma^2 \mathbf{I}_{ST \times ST} \quad (4)$$

where  $\mathbf{I}_{ST \times ST}$  is an  $ST \times ST$  identity matrix. Hence,  $|\mathbf{\Lambda}| = \sigma^{2ST}$ . Also define

$$\mathbf{a} = [C_1, G_1, C_2, G_2, \dots, C_k, G_k]^T. \quad (5)$$

Let

$$\mathbf{A}_i = \begin{bmatrix} e^{j[0\omega_i + 0\nu_i]}, \dots, e^{j[0\omega_i + (T-1)\nu_i]}, \dots, \dots, e^{j[(S-1)\omega_i + (T-1)\nu_i]} \end{bmatrix}^T \quad (6)$$

and let us define the following  $ST \times 2k$  matrix

$$\mathbf{D} = [\text{Re}(A_1), \text{Im}(A_1), \text{Re}(A_2), \text{Im}(A_2), \dots, \dots, \text{Re}(A_k), \text{Im}(A_k)] \quad (7)$$

Using the foregoing notations we have that

$$\mathbf{y} = \mathbf{D}\mathbf{a} + \mathbf{u}. \quad (8)$$

In the following it is assumed that the matrix  $\mathbf{D}^T \mathbf{D}$  is full rank.

## 3. MAP MODEL ORDER SELECTION CRITERION

Let  $p(k)$  be the *a-priori* probability that there exist  $k$  sinusoidal components in the observed field. It is assumed that there are  $Q$  competing models, where  $Q > M$  ( $M$  being the actual number of sinusoidal components), and that each model is equiprobable. That is

$$p(k) = \frac{1}{Q}, \quad k \in Z_Q \quad (9)$$

where  $Z_Q = \{0, 1, 2, \dots, Q\}$ . The MAP estimate of  $M$  is the value of  $k$  that maximizes the *a-posteriori* probability  $p(k|\mathbf{y})$ , where  $k \in Z_Q$ . More specifically,

$$\begin{aligned} \hat{M}_{MAP} &= \arg \max_{k \in Z_Q} \{p(k|\mathbf{y})\} \\ &= \arg \max_{k \in Z_Q} \left\{ \frac{p(\mathbf{y}|k)p(k)}{p(\mathbf{y})} \right\} \\ &= \arg \max_{k \in Z_Q} \{p(\mathbf{y}|k)\} \\ &= \arg \max_{k \in Z_Q} \left\{ \ln p(\mathbf{y}|k) \right\} \end{aligned} \quad (10)$$

where  $p(\mathbf{y}|k)$  denotes the conditional probability of  $\mathbf{y}$  given that there are  $k$  sinusoidal components in the data.

Let

$$\mathbf{W} = [\omega_1, \omega_2, \dots, \omega_k, \nu_1, \nu_2, \dots, \nu_k]^T. \quad (11)$$

Also let  $\mathcal{R}^+$  denote the positive real line, let  $\mathcal{A}_k = \mathcal{R}^{2k}$ , and let  $\Omega_k = ([0, 2\pi])^{2k}$ . Thus, we have that  $\sigma \in \mathcal{R}^+$ ,  $\mathbf{a} \in \mathcal{A}_k$ , and  $\mathbf{W} \in \Omega_k$ . Using these notations the conditional probability density  $p(\mathbf{y}|k)$  is expressed by

$$p(\mathbf{y}|k) = \int_{\Omega_k} \int_{\mathcal{R}^+} \int_{\mathcal{A}_k} p(\mathbf{y}|k, \mathbf{W}, \sigma, \mathbf{a}) \times p(\mathbf{W}, \sigma, \mathbf{a}|k) d\mathbf{a} d\sigma d\mathbf{W} \quad (12)$$

where  $p(\mathbf{W}, \sigma, \mathbf{a}|k)$  is the *a-priori* probability of  $\mathbf{W}$ ,  $\sigma$  and  $\mathbf{a}$  given there exist  $k$  sinusoidal components in the observed data.

## 4. DERIVATION OF THE CRITERION

### 4.1. Priors Selection

Inspecting (10) and (12) we conclude that finding  $\hat{M}_{MAP}$  using the observed data only, requires that some assumptions be made regarding the prior distribution of the model parameters,  $p(\mathbf{W}, \sigma, \mathbf{a}|k)$ . Clearly our goal is to derive a model selection rule that will be based on a non-informative prior about the parameters. In other words, the selected prior should be chosen such that it represents the lack of *a-priori* knowledge of the values of problem parameters, before the data is observed. (See, *e.g.*, [2] for a detailed discussion of the problem of choosing non-informative priors).

Clearly,

$$p(\mathbf{W}, \sigma, \mathbf{a}|k) = p(\sigma, \mathbf{a}|\mathbf{W}, k)p(\mathbf{W}|k). \quad (13)$$

Since the sinusoidal frequencies are assumed independent of each other (*i.e.*, that they are not harmonically related), the lack of *a-priori* knowledge of the frequencies is modeled by assuming the frequencies  $(\omega_i, \nu_i)$  to be uniformly distributed on  $\Omega_k$ . Thus,

$$p(\mathbf{W}|k) = \frac{1}{(2\pi)^{2k}}. \quad (14)$$

Note that since the probability of  $\omega_i$  being equal to  $\omega_j$  for some  $i \neq j$  is zero (and similarly for  $\nu_i$  being equal to  $\nu_j$ ), we assume in the following that for all  $i \neq j$ ,  $\omega_i \neq \omega_j$  (and similarly  $\nu_i \neq \nu_j$ ). Hence the following derivation of the model order selection criterion holds almost everywhere in the problem probability space,

*i.e.*, except for a set of models of probability measure zero.

Given that  $\mathbf{W}$  and  $k$  are known,  $\mathbf{D}$  is also known and the observation model (8) becomes a linear regression model where the observations are subject to a zero mean white Gaussian observation noise with variance  $\sigma^2$ , such that  $\mathbf{a}$ ,  $\sigma$  are unknown. For this problem it is shown in [2] that in the space defined by  $\mathbf{a}$  and  $\ln \sigma$  the shape of the likelihood function surface is “data translated”, *i.e.*, it is invariant to translations that result from the different values these parameters assume in different realizations of the observed data. Hence the idea that little is known *a-priori* relative to the information contained in the observed data is expressed by choosing a prior distribution such that  $p(\ln \sigma, \mathbf{a}|\mathbf{W}, k)$  is locally uniform, or equivalently that

$$p(\sigma, \mathbf{a}|\mathbf{W}, k) \propto \sigma^{-1}. \quad (15)$$

Substituting (14) and (15) into (13) we have that the desired non-informative prior is given by

$$p(\mathbf{W}, \sigma, \mathbf{a}|k) \propto \frac{1}{(2\pi)^{2k}} \sigma^{-1}. \quad (16)$$

### 4.2. Evaluation of the a-Posteriori Distribution

In this subsection we derive an approximate expression for the *a-posteriori* probability distribution  $p(\mathbf{y}|k)$  given in (12). Since the noise field  $\{u(n, m)\}$  is Gaussian we have using (4) and (8)

$$p(\mathbf{y}|k, \mathbf{W}, \sigma, \mathbf{a}) = p(\mathbf{u}|\sigma) = (2\pi\sigma^2)^{-\frac{ST}{2}} \exp\left\{-\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{D}\mathbf{a})^T(\mathbf{y} - \mathbf{D}\mathbf{a})\right\}. \quad (17)$$

Let  $\hat{\mathbf{a}} = (\mathbf{D}^T \mathbf{D})^{-1} \mathbf{D}^T \mathbf{y}$  and let  $\mathbf{P}^\perp$  denote the projection matrix defined by

$$\mathbf{P}^\perp = \mathbf{I} - \mathbf{D}(\mathbf{D}^T \mathbf{D})^{-1} \mathbf{D}^T. \quad (18)$$

Using these notations we have that

$$(\mathbf{y} - \mathbf{D}\mathbf{a})^T(\mathbf{y} - \mathbf{D}\mathbf{a}) = \mathbf{y}^T \mathbf{P}^\perp \mathbf{y} + (\mathbf{a} - \hat{\mathbf{a}})^T \mathbf{D}^T \mathbf{D} (\mathbf{a} - \hat{\mathbf{a}}). \quad (19)$$

Applying the prior (16) and evaluating the marginal distribution we have

$$\begin{aligned} p(\mathbf{y}, \mathbf{W}, \sigma|k) &= \int_{\mathcal{A}_k} p(\mathbf{y}|k, \mathbf{W}, \sigma, \mathbf{a}) p(\mathbf{W}, \sigma, \mathbf{a}|k) d\mathbf{a} \\ &\propto \int_{\mathcal{A}_k} (2\pi\sigma^2)^{-\frac{ST}{2}} \exp\left\{-\frac{1}{2\sigma^2}(\mathbf{y} - \mathbf{D}\mathbf{a})^T(\mathbf{y} - \mathbf{D}\mathbf{a})\right\} \\ &\quad \times \frac{1}{(2\pi)^{2k}\sigma} d\mathbf{a} \\ &= (2\pi\sigma^2)^{-\frac{ST}{2}} \frac{1}{(2\pi)^{2k}\sigma} \exp\left\{-\frac{1}{2\sigma^2} \mathbf{y}^T \mathbf{P}^\perp \mathbf{y}\right\} \end{aligned}$$

$$\begin{aligned} & \times \int_{\mathcal{A}_k} \exp \left\{ -\frac{1}{2\sigma^2} (\mathbf{a} - \hat{\mathbf{a}})^T \mathbf{D}^T \mathbf{D} (\mathbf{a} - \hat{\mathbf{a}}) \right\} d\mathbf{a} \\ & = (2\pi\sigma^2)^{-\frac{ST}{2}} \frac{1}{(2\pi)^{2k}\sigma} \exp \left\{ -\frac{1}{2\sigma^2} \mathbf{y}^T \mathbf{P}^\perp \mathbf{y} \right\} \frac{(\sqrt{2\pi}\sigma)^{2k}}{|\mathbf{D}^T \mathbf{D}|^{1/2}}. \end{aligned} \quad (20)$$

Next, we evaluate  $p(\mathbf{y}, \mathbf{W}|k)$ . Substituting (20) we have

$$\begin{aligned} p(\mathbf{y}, \mathbf{W}|k) &= \int_{\mathcal{R}^+} p(\mathbf{y}, \mathbf{W}, \sigma|k) d\sigma \\ &\propto 2^{-2k-1} \pi^{-\frac{ST+2k}{2}} \left( \frac{ST-2k}{2} \right) |\mathbf{D}^T \mathbf{D}|^{-\frac{1}{2}} (\mathbf{y}^T \mathbf{P}^\perp \mathbf{y})^{-\frac{ST-2k}{2}} \end{aligned} \quad (21)$$

where  $(\cdot)$  is the standard Gamma function (see, *e.g.*, [2] for the integration result).

Finally, to obtain an expression for the conditional probability  $p(\mathbf{y}|k)$  we have to evaluate

$$p(\mathbf{y}|k) = \int_{\Omega_k} p(\mathbf{y}, \mathbf{W}|k) d\mathbf{W}. \quad (22)$$

Since a direct analytic solution to this integration problem does not exist, we derive an approximate solution, employing the Laplace integration method (see, *e.g.*, [3]). Following [3], p. 71, we first expand  $\frac{1}{ST} \ln p(\mathbf{y}, \mathbf{W}|k)$  into a Taylor series about  $\hat{\mathbf{W}}$ , where  $\hat{\mathbf{W}}$  denotes the ML estimate of  $\mathbf{W}$ . Since  $\hat{\mathbf{W}}$  is a maximum point of the likelihood function, the first order derivatives of  $\frac{1}{ST} \ln p(\mathbf{y}, \mathbf{W}|k)$  at this point vanish. Omitting from the expansion terms of order higher than two, we have

$$\begin{aligned} p(\mathbf{y}, \mathbf{W}|k) &= \exp \left\{ ST \frac{\ln p(\mathbf{y}, \mathbf{W}|k)}{ST} \right\} \\ &\simeq \exp \left\{ ST \frac{\ln p(\mathbf{y}, \hat{\mathbf{W}}|k)}{ST} - \frac{ST}{2} (\mathbf{W} - \hat{\mathbf{W}})^T \hat{\mathbf{H}}_{ML} (\mathbf{W} - \hat{\mathbf{W}}) \right\} \end{aligned} \quad (23)$$

where

$$\begin{aligned} \mathbf{H}_{ML} &= \\ & -\frac{1}{ST} \begin{bmatrix} \frac{\partial^2 \ln p(\mathbf{y}, \mathbf{W}|k)}{\partial \mathbf{W}_1^2} & \frac{\partial^2 \ln p(\mathbf{y}, \mathbf{W}|k)}{\partial \mathbf{W}_1 \partial \mathbf{W}_2} & \dots & \frac{\partial^2 \ln p(\mathbf{y}, \mathbf{W}|k)}{\partial \mathbf{W}_1 \partial \mathbf{W}_{2k}} \\ \frac{\partial^2 \ln p(\mathbf{y}, \mathbf{W}|k)}{\partial \mathbf{W}_2 \partial \mathbf{W}_1} & \frac{\partial^2 \ln p(\mathbf{y}, \mathbf{W}|k)}{\partial \mathbf{W}_2^2} & \dots & \frac{\partial^2 \ln p(\mathbf{y}, \mathbf{W}|k)}{\partial \mathbf{W}_2 \partial \mathbf{W}_{2k}} \\ \vdots & \vdots & \dots & \vdots \\ \frac{\partial^2 \ln p(\mathbf{y}, \mathbf{W}|k)}{\partial \mathbf{W}_{2k} \partial \mathbf{W}_1} & \frac{\partial^2 \ln p(\mathbf{y}, \mathbf{W}|k)}{\partial \mathbf{W}_{2k} \partial \mathbf{W}_2} & \dots & \frac{\partial^2 \ln p(\mathbf{y}, \mathbf{W}|k)}{\partial \mathbf{W}_{2k}^2} \end{bmatrix} \end{aligned} \quad (24)$$

is the Hessian matrix of  $\frac{1}{ST} \ln p(\mathbf{y}, \mathbf{W}|k)$  evaluated at  $\mathbf{W} = \hat{\mathbf{W}}$ . As  $\hat{\mathbf{W}}$  is a maximum point of  $\ln p(\mathbf{y}, \mathbf{W}|k)$ ,  $\mathbf{H}_{ML}$  is positive definite. Since  $\ln p(\mathbf{y}, \mathbf{W}|k)$  is assumed sufficiently smooth at  $\hat{\mathbf{W}}$ ,  $\mathbf{H}_{ML}$  is symmetric.

Substituting (23) into (22) and employing the Laplace asymptotic approximation we have that as  $ST \rightarrow \infty$

$$\begin{aligned} p(\mathbf{y}|k) &= \int_{\Omega_k} \exp \left\{ ST \frac{\ln p(\mathbf{y}, \mathbf{W}|k)}{ST} \right\} d\mathbf{W} \\ &\approx p(\mathbf{y}, \hat{\mathbf{W}}|k) (2\pi)^k |\hat{\mathbf{H}}_{ML}|^{-\frac{1}{2}} (ST)^{-k} \end{aligned} \quad (25)$$

**Lemma 1**

$$|\hat{\mathbf{H}}_{ML}| = O(S^{2k} T^{2k}). \quad (26)$$

*Proof:* See [9].

Substituting (21) and (26) into (25) we have

$$\begin{aligned} p(\mathbf{y}|k) &\propto 2^{-k-1} \pi^{-\frac{ST}{2}} \left( \frac{ST-2k}{2} \right) |\hat{\mathbf{D}}^T \hat{\mathbf{D}}|^{-\frac{1}{2}} \\ &\quad \times \left( \mathbf{y}^T \hat{\mathbf{P}}^\perp \mathbf{y} \right)^{-\frac{ST-2k}{2}} O(S^{-2k} T^{-2k}) \end{aligned} \quad (27)$$

where  $\hat{\mathbf{D}}$  and  $\hat{\mathbf{P}}^\perp$  are the matrices  $\mathbf{D}$  and  $\mathbf{P}^\perp$ , respectively, with  $\mathbf{W}$  substituted by its ML estimate,  $\hat{\mathbf{W}}$ . It is possible to further simplify (27) by observing that  $|\mathbf{D}^T \mathbf{D}| = O(S^{2k} T^{2k})$  (see [9]). Furthermore, employing the asymptotic properties of the Gamma function (see, *e.g.*, [8], p. 31) we have that as  $ST \rightarrow \infty$ ,

$$\left( \frac{ST-2k}{2} \right) = \frac{ST-2k-1}{2} \ln \left( \frac{ST}{2} \right) - \frac{ST}{2} + O(1). \quad (28)$$

Substituting these approximations into (27), and omitting terms that are independent of  $k$ , the final form of the model order selection criterion can be readily established:

$$\begin{aligned} \hat{M}_{MAP} &= \arg \min_{k \in Z_Q} \left\{ -\ln p(\mathbf{y}|k) \right\} \\ &= \arg \min_{k \in Z_Q} \left\{ \frac{ST-2k}{2} \ln(\mathbf{y}^T \hat{\mathbf{P}}^\perp \mathbf{y}) + \frac{1}{2} \ln |\hat{\mathbf{D}}^T \hat{\mathbf{D}}| \right. \\ &\quad \left. + k \ln \frac{ST}{2} + 2k \ln ST + (k+1) \ln 2 \right\} \\ &= \arg \min_{k \in Z_Q} \left\{ \frac{ST-2k}{2} \ln(\mathbf{y}^T \hat{\mathbf{P}}^\perp \mathbf{y}) + 4k \ln ST \right\} \end{aligned} \quad (29)$$

## 5. NUMERICAL RESULTS

To illustrate the performance of the proposed model order selection rule we present some numerical examples. In the examples below, the data field was generated with four equiamplitude sinusoidal components, and we define

$$\text{SNR}_i = 10 \log \frac{C_i^2 + G_i^2}{2\sigma^2} \quad (30)$$

The noise is a white Gaussian noise field with variance  $\sigma^2$  which is chosen to yield the desired signal to noise ratio. In these experiments the signal to noise ratio of each component,  $\text{SNR}_i$ , varies in the range of -15dB to -5dB, in steps of 1dB. For each SNR, 100 Monte-Carlo experiments are performed. The data field dimensions are  $32 \times 32$ . The frequencies of the sinusoidal components are  $(-2\pi 0.155, 2\pi 0.253)$ ,  $(-2\pi 0.155, 2\pi 0.296)$ ,  $(-2\pi 0.112, 2\pi 0.274)$ ,  $(2\pi 0.112, 2\pi 0.201)$ . Their amplitudes are given by  $C_i = G_i = 1, i = 1, \dots, 4$ . The performance results of the proposed MAP selection criterion are summarized in Table 1 for various values of  $\text{SNR}_i$ . For comparison, the performance results of the GAIC criterion, [6], are listed as well. To further illustrate the performance of the proposed MAP model order selection criterion, the probabilities of correct model order selection for the two criteria are depicted in Fig. 1. The simulation results demonstrate that even for modest dimensions of the observed field, and relatively low SNR's, *i.e.*, as low as -9dB, the error rates of both the MAP and the GAIC model order selection criteria are very low. The performance of the MAP rule is shown to be better than that of the GAIC for lower SNR's. Furthermore, the results indicate that for the lower SNR range, the probability of correct model order selection by the MAP criterion is not only higher, but also that the magnitude of the error is much smaller than in the case of the GAIC model order estimate.

$\text{SNR}_i$		k=1	k=2	k=3	k=4
-15dB	MAP	29	34	29	8
	GAIC	94	6	0	0
-14dB	MAP	4	27	45	24
	GAIC	86	12	2	0
-13dB	MAP	3	13	46	38
	GAIC	57	33	8	2
-12dB	MAP	0	3	18	79
	GAIC	22	23	27	28
-11dB	MAP	0	0	7	93
	GAIC	2	6	21	71
-10dB	MAP	0	0	4	96
	GAIC	0	0	8	92
-9dB	MAP	0	0	0	100
	GAIC	0	0	0	100

Table 1: Performance comparison of MAP and GAIC criteria for various values of  $\text{SNR}_i$ .

## 6. REFERENCES

- [1] P. M. Djuric, "A Model Selection Rule for Sinusoids in White Gaussian Noise," *IEEE Trans. Signal Process.*, vol. 44, pp. 1744-1751, 1996.
- [2] G. E. P. Box and G. C. Tiao, *Bayesian Inference in Statistical Analysis.*, New York: Wiley, 1992.
- [3] N. G. De Bruijn, *Asymptotic Methods in Analysis*, 3rd edition, Amsterdam: North-Holland Publishing Co., 1970.
- [4] L. Kavalieris and E. J. Hannan, "Determining the Number of Terms in a Trigonometric Regression," *J. Time Series Anal.*, vol. 15, pp. 613-625, 1994.
- [5] P. Stoica, P. Eykhoff, P. Janssen and T. Soderstrom, "Model-Structure Selection by Cross-Validation," *Int. J. Control*, vol. 43, pp. 1841-1878, 1986.
- [6] J. Li and P. Stoica, "Efficient Mixed-Spectrum Estimation with Application to Target Feature Extraction," *IEEE Trans. Signal Process.*, vol. 44, pp. 281-295, 1996.
- [7] R. Stoica, J. Zerubia and J. M. Francos, "The Two-Dimensional Wold Decomposition for Segmentation and Indexing in Image Libraries," *Int. Conf. Acoust., Speech, Signal Processing*, Seattle, 1998.
- [8] E. D. Rainville, *Special Functions*, MacMillan, New York, 1967.
- [9] M. Kliger and J. M. Francos, "MAP Model Order Selection Criterion for 2-D Sinusoids in Noise," in preparation.

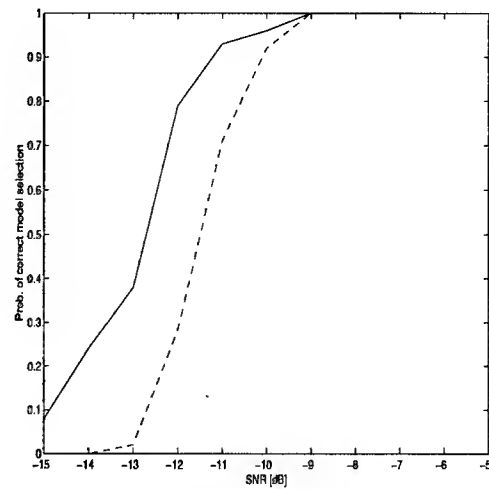


Figure 1: Probabilities of correct model order selection. The solid and the dashed lines represent the MAP and the GAIC performance curves, respectively.

# OPTIMUM LINEAR PERIODICALLY TIME-VARYING FILTER

Dong Wei

Center for Telecommunications and Information Networking  
Department of Electrical and Computer Engineering, Drexel University  
Philadelphia, PA 19104 U.S.A.  
E-mails: wei@ece.drexel.edu

## ABSTRACT

We study the optimum (in the minimum mean-square error sense) linear periodically time-varying deconvolution filter of finite size. We show that the filter can be in the form of lapped transform or multirate filterbank, and it includes the FIR Wiener filter as a special case. We demonstrate that the proposed filter always possesses a gain over the Wiener filter.

## 1. INTRODUCTION

Consider the discrete-time model

$$x[n] = (s * h)[n] + w[n] \quad (1)$$

$$= \sum_{m=0}^{N-1} h[m]s[n-m] + w[n] \quad (2)$$

where  $s[n]$  is the original signal,  $h[n]$  is a known linear time-invariant (LTI) system with  $N$  taps,  $w[n]$  is the additive noise,  $x[n]$  is the observed data, and the symbol  $*$  denotes convolution. We assume that both  $s[n]$  and  $w[n]$  are zero-mean, wide-sense stationary, second-order random processes, their second-order statistics are known, and they are uncorrelated, i.e.,

$$E\{s[n]w^*[k]\} = E\{s[n]\}E\{w^*[k]\} \quad (3)$$

for any  $n$  and  $k$ . Such a model has been widely used in signal processing applications such as filtering, smoothing, prediction, noise canceling, and deconvolution, just to name a few.

The goal is to estimate the signal  $s[n]$  from the noisy, filtered data  $x[n]$ . An LTI finite impulse response (FIR) filter  $f[n]$  can be applied to  $x[n]$ . The resulting estimate of  $s[n]$  is given by

$$\hat{s}[n] = (x * f)[n] \quad (4)$$

$$= \sum_{m=0}^{K-1} f[m]x[n-m] \quad (5)$$

where  $K$  is the length of  $f[n]$ . The FIR Wiener deconvolution filter is the optimum LTI FIR system (denoted by the vector  $\mathbf{f}_{\text{opt}}$ ) in the minimum mean-square error (MMSE) sense:

$$\mathbf{f}_{\text{opt}} = \arg \min_{\mathbf{f}} E\{|\hat{s}[n] - s[n]|^2\} \quad (6)$$

where

$$\mathbf{f} = [f[0] \ f[1] \ \dots \ f[K-1]]^T \quad (7)$$

with the symbol  $^T$  denoting matrix transpose.

We now reconsider the optimality of the Wiener filter in (6) from a different viewpoint. The filtering operation in (4) can be expressed as

$$\hat{\mathbf{s}}[n] = \mathbf{F}_{\text{LTI}} \mathbf{x}[n] \quad (8)$$

where

$$\hat{\mathbf{s}}[n] = [\hat{s}[n] \ \hat{s}[n-1] \ \dots \ \hat{s}[n-M+1]]^T, \quad (9)$$

$$\mathbf{x}[n] = [x[n] \ x[n-1] \ \dots \ x[n-L+1]]^T, \quad (10)$$

$\mathbf{F}_{\text{LTI}}$  is an  $M \times L$  matrix given by

$$\mathbf{F}_{\text{LTI}} = \begin{bmatrix} \mathbf{f}^T & 0 & 0 & \dots & 0 \\ 0 & \mathbf{f}^T & 0 & \dots & 0 \\ \vdots & & & & \\ 0 & 0 & \dots & 0 & \mathbf{f}^T \end{bmatrix}, \quad (11)$$

and  $L = K + M - 1$ .

The linear lapped transform [1]

$$\hat{\mathbf{s}}[n] = \mathbf{F} \mathbf{x}[n] \quad (12)$$

where  $\mathbf{F}$  is an  $M \times L$  constant matrix, is a more general version of linear filtering than (8). We require that  $M \leq L$ . When  $M = L$ , the linear lapped transform reduces to a linear block transform.

A few interesting questions arise. Does the Wiener filter result in the optimum (in the MMSE sense) estimate  $\hat{\mathbf{s}}[n]$ ? If it does not, how can we do better and what is the best estimate?

In this paper, we answer these questions.

## 2. MINIMUM MEAN SQUARE ERROR LINEAR PERIODICALLY TIME-VARYING FILTERING

### 2.1. Some Basics

Since

$$\hat{s}[n - lM] = \mathbf{F}\mathbf{x}[n - lM] \quad (13)$$

for any integer  $l$ , such a linear lapped transform is in general a generic LPTV filter with period  $M$ . The LPTV filter can be implemented by means of an  $M$ -channel multirate filterbank [2].

When  $M = 1$ , the LPTV filter reduces to an LTI filter with  $L$  taps. For  $M > 1$ , the LPTV filter reduces to an LTI filter if and only if the  $M \times L$  matrix  $\mathbf{F}$  satisfies. This implies that any LTI filter of length up to  $L - M + 1$  is a special case of the LPTV filter characterized by  $\mathbf{F}$ . Therefore, the optimum LPTV filter of size  $M \times L$  always possesses a gain over the Wiener filter of length  $L - M + 1$ . Such a gain results from the more flexible processing of data blocks than the LTI filtering. For filtering, the two filters require  $L$  and  $L - M + 1$  multiplications per data sample, respectively. When  $M$  is small compared to  $L$ , their computational complexities are comparable.

### 2.2. The Optimum Filter

The model in (1) can be expressed in the vector form:

$$\mathbf{x}[n] = \mathbf{H}\mathbf{s}[n] + \mathbf{w}[n] \quad (14)$$

where  $\mathbf{H}$  is an  $L \times (L + N - 1)$  matrix given by

$$\mathbf{H} = \begin{bmatrix} \mathbf{h}^T & 0 & 0 & \dots & 0 \\ 0 & \mathbf{h}^T & 0 & \dots & 0 \\ \vdots & & & & \\ 0 & 0 & \dots & 0 & \mathbf{h}^T \end{bmatrix}, \quad (15)$$

$$\mathbf{h} = [h[0] \ h[1] \ \dots \ h[N - 1]]^T, \quad (16)$$

$$\mathbf{s}[n] = [s[n] \ s[n - 1] \ \dots \ s[n - L - N + 2]]^T, \quad (17)$$

and

$$\mathbf{w}[n] = [w[n] \ w[n - 1] \ \dots \ w[n - L + 1]]^T. \quad (18)$$

Define the estimation error in the  $n$ th block as

$$\mathbf{e}[n] = \hat{\mathbf{s}}[n] - \mathbf{A}\mathbf{s}[n] \quad (19)$$

where

$$\mathbf{A} = [\mathbf{I}_M \ \mathbf{0}_{M \times (L - M + N - 1)}], \quad (20)$$

and

$$\mathbf{e}[n] = \mathbf{F}(\mathbf{H}\mathbf{s}[n] + \mathbf{w}[n]) - \mathbf{A}\mathbf{s}[n] \quad (21)$$

$$= (\mathbf{F}\mathbf{H} - \mathbf{A})\mathbf{s}[n] + \mathbf{F}\mathbf{w}[n]. \quad (22)$$

We attempt to design the optimum  $\mathbf{F}$  to minimize the mean-square error (MSE) in the block  $\hat{\mathbf{s}}[n]$ , which is given by

$$\mathcal{J} = \frac{1}{M} E\{\mathbf{e}^H[n]\mathbf{e}[n]\}. \quad (23)$$

It follows that

$$\mathcal{J} = \frac{1}{M} E\{\|(\mathbf{F}\mathbf{H} - \mathbf{A})\mathbf{s}[n] + \mathbf{F}\mathbf{w}[n]\|^2\} \quad (24)$$

$$= \frac{1}{M} E\{\text{tr}\{((\mathbf{F}\mathbf{H} - \mathbf{A})\mathbf{s}[n] + \mathbf{F}\mathbf{w}[n]) \times ((\mathbf{F}\mathbf{H} - \mathbf{A})\mathbf{s}[n] + \mathbf{F}\mathbf{w}[n])^H)\}\} \quad (25)$$

$$= \frac{1}{M} \text{tr}[(\mathbf{F}\mathbf{H} - \mathbf{A})\mathbf{R}_s(\mathbf{F}\mathbf{H} - \mathbf{A})^H + \mathbf{F}\mathbf{R}_w\mathbf{F}^H] \quad (26)$$

$$= \frac{1}{M} \text{tr}[\mathbf{F}(\mathbf{H}\mathbf{R}_s\mathbf{H}^H + \mathbf{R}_w)\mathbf{F}^H + \mathbf{A}\mathbf{R}_s\mathbf{A}^H - \mathbf{F}\mathbf{H}\mathbf{R}_s\mathbf{A}^H - \mathbf{A}\mathbf{R}_s\mathbf{H}^H\mathbf{F}^H] \quad (27)$$

where

$$\mathbf{R}_s = E\{\mathbf{s}[n]\mathbf{s}^H[n]\}, \quad (28)$$

$$\mathbf{R}_w = E\{\mathbf{w}[n]\mathbf{w}^H[n]\}. \quad (29)$$

Setting

$$\left. \frac{\partial \mathcal{J}}{\partial \mathbf{F}} \right|_{\mathbf{F}=\mathbf{F}_{\text{opt}}} = \mathbf{0}_{M \times L}, \quad (30)$$

we obtain the matrix form of the Wiener-Hopf equations:

$$\mathbf{F}_{\text{opt}}(\mathbf{H}\mathbf{R}_s\mathbf{H}^H + \mathbf{R}_w) = \mathbf{A}\mathbf{R}_s\mathbf{H}^H. \quad (31)$$

Therefore, the optimum LPTV filter is

$$\mathbf{F}_{\text{opt}} = \mathbf{A}\mathbf{R}_s\mathbf{H}^H(\mathbf{H}\mathbf{R}_s\mathbf{H}^H + \mathbf{R}_w)^{-1} \quad (32)$$

and the resulting minimum MSE is

$$\mathcal{J}_{\text{LPTV,min}} = \sigma_s^2 - \frac{1}{M} \text{tr}[\mathbf{A}\mathbf{R}_s\mathbf{H}^H(\mathbf{H}\mathbf{R}_s\mathbf{H}^H + \mathbf{R}_w)^{-1} \times \mathbf{H}\mathbf{R}_s\mathbf{A}^H]. \quad (33)$$

The optimum filter can be viewed as the extension of the FIR Wiener filter to LPTV system. Indeed, when  $M = 1$ ,  $\mathbf{F}_{\text{opt}}$  reduces to the FIR Wiener filter with  $L$  taps. On the other hand, for  $D = 1, 2, \dots, M$ , the  $D$ th row of the filtering matrix  $\mathbf{F}_{\text{opt}}$  is the MMSE FIR filter for estimating  $s[n - D + 1]$  from the data set  $\{x[m] : -\infty < m \leq n\}$ .

### 2.3. When Is There No Gain?

In general, the performance of the optimum LPTV filter is better than the performance of the optimum LTI filter in the sense that

$$\mathcal{J}_{\text{LPTV,min}} \leq \mathcal{J}_{\text{LTI,min}} \quad (34)$$

where the equality holds if and only if



- the signal  $s[n]$  is a white noise process, i.e.,

$$E\{s[n]s^*[n+l]\} = \sigma_s^2 \delta[l], \quad (35)$$

- the noise  $w[n]$  is a white noise process, i.e.,

$$E\{w[n]w^*[n+l]\} = \sigma_w^2 \delta[l], \quad (36)$$

and

- the LTI system  $h[n]$  has only one tap, i.e.,  $N = 1$ .

#### 2.4. Asymptotic Performance Analysis

We assume that  $s[n]$  and  $w[n]$  are both regular processes with rational power spectra.

Let  $f_D[n]$  denote the causal, infinite impulse response (IIR) Wiener deconvolution filter for estimating  $s[n-D]$  from the data set  $\{x[m] : -\infty < m \leq n\}$ , where  $D \geq 0$  indicates a delay. The performance of  $f_D[n]$  shall be used in our analysis of the asymptotic performance of the proposed optimum LPTV filter. The transfer function of  $f_D[n]$  is given by

$$F_D(z) = \frac{1}{\sigma_0^2 Q(z)} \left[ \frac{z^{-D} P_s(z) H^*(1/z^*)}{Q^*(1/z^*)} \right]_+ \quad (37)$$

where  $Q(z)$  is the monic, minimum-phase factor determined by the spectral factorization of the power spectrum of  $x[n]$ :

$$P_x(z) = H(z)H^*(1/z^*)P_s(z) + P_w(z) \quad (38)$$

$$= \sigma_0^2 Q(z)Q^*(1/z^*) \quad (39)$$

and the subscript “+” is used to indicate the “positive-time part” of the sequence whose  $z$ -transform is contained within the brackets. The resulting MSE is

$$\mathcal{J}_{\text{IIR},\min}^{(D)} = r_s(0) - \sum_{l=0}^{\infty} \sum_{n=0}^{N-1} f_D[l]h[n]r_s[D-n-l] \quad (40)$$

or

$$\mathcal{J}_{\text{IIR},\min}^{(D)} = \frac{1}{2\pi} \int_{-\pi}^{\pi} P_s(e^{j\omega}) \times [1 - F_D(e^{j\omega})H(e^{j\omega})e^{j\omega D}] d\omega. \quad (41)$$

The derivations of the optimum filter  $f_D[n]$  and its associated MSE are given in Appendix.

In general, increasing the delay  $D$  leads to the smaller  $\mathcal{J}_{\text{IIR},\min}^{(D)}$ . Asymptotically,  $f_{\infty}[n]$  is the non-causal IIR Wiener filter.

The performance of estimating the signal block  $\mathbf{A}s[n]$  using the filtering matrix  $\mathbf{F}$  can be improved if more observed data are processed, or equivalently, the parameter  $L$  is increased. As  $L$  tends to infinity, the

MMSE FIR filter converges to the MMSE causal IIR filter. Therefore,

$$\mathcal{J}_{\text{FIR},\min} > \lim_{L \rightarrow \infty} \mathcal{J}_{\text{FIR},\min} \quad (42)$$

$$= \mathcal{J}_{\text{IIR},\min}^{(0)}. \quad (43)$$

If  $M$  is fixed and  $L$  tends to infinity, then the  $D$ th row of the optimum filtering matrix  $\mathbf{F}_{\text{opt}}$  converges to  $f_D[n]$ . Therefore,

$$\mathcal{J}_{\text{LPTV},\min} > \lim_{L \rightarrow \infty} \mathcal{J}_{\text{LPTV},\min} \quad (44)$$

$$= \frac{1}{M} \sum_{D=0}^{M-1} \mathcal{J}_{\text{IIR},\min}^{(D)}. \quad (45)$$

If both  $M$  and  $L$  approach to infinity with  $K = L - M + 1$  fixed, then

$$\lim_{L \rightarrow \infty, M \rightarrow \infty} \mathcal{J}_{\text{LPTV},\min} = \lim_{M \rightarrow \infty} \frac{1}{M} \sum_{D=0}^{M-1} \mathcal{J}_{\text{IIR},\min}^{(D)} \quad (46)$$

$$= \mathcal{J}_{\text{IIR},\min}^{(\infty)} \quad (47)$$

which corresponds to the MSE of the non-causal IIR Wiener filter.

In summary, the optimum LPTV filter outperforms the Wiener filter asymptotically.

### 3. CONCLUSION

We have presented the MMSE linear periodically time-varying deconvolution filter. The proposed filter outperforms the linear time-invariant counterpart at the expense of increase in computational complexity and delay.

### APPENDIX

We now prove (37) and (40).

According to the model given in (1), we first whiten the process  $x[n]$  to obtain a unit-variance white noise process:

$$y[n] = (b * x)[n] \quad (48)$$

where the whitening filter  $b[n]$  is given by

$$B(z) = \frac{1}{\sigma_0 Q(z)} \quad (49)$$

which is causal and stable. Next, we obtain the estimate of  $s[n-D]$  by filtering  $y[n]$  with a causal IIR filter  $g[n]$ :

$$\hat{s}[n-D] = \sum_{m=0}^{\infty} g[m]y[n-m]. \quad (50)$$

To minimize  $E\{|s[n-D] - \hat{s}[n-D]|^2\}$  with respect to  $g[n]$ , we use the orthogonality principle to obtain the Wiener-Hopf equations

$$E\{(s[n-D] - \hat{s}[n-D])y^*[n-k]\} = 0 \quad (51)$$

for  $0 \leq k < \infty$ , or equivalently,

$$r_{sy}[k-D] = \sum_{m=0}^{\infty} g[m]r_y[k-m] \quad (52)$$

$$= g[k] \quad (53)$$

for  $0 \leq k < \infty$ . Therefore,

$$G(z) = [z^{-D}P_{sy}(z)]_+. \quad (54)$$

Since

$$r_{sy}[k] = E\{s[n]y^*[n-k]\} \quad (55)$$

$$= E\left\{s[n] \left( \sum_{l=0}^{\infty} b[l]x[n-k-l] \right)^*\right\} \quad (56)$$

$$= \sum_{l=0}^{\infty} b^*[l]r_{sx}[k+l] \quad (57)$$

$$= \sum_{l=0}^{\infty} b^*[l]E\{s[n+k+l] \times \left( \sum_{m=0}^{N-1} h[m]s[n-m] + w[n] \right)^*\} \quad (58)$$

$$= \sum_{l=0}^{\infty} \sum_{m=0}^{N-1} b^*[l]h^*[m]r_s[k+l+m] \quad (59)$$

which implies that

$$P_{sy}(z) = B^*(1/z^*)H^*(1/z^*)P_s(z) \quad (60)$$

$$= \frac{H^*(1/z^*)P_s(z)}{\sigma_0 Q^*(1/z^*)}. \quad (61)$$

Since the causal, IIR Wiener deconvolution filter for estimating  $s[n-D]$  from the data set  $\{x[m] : -\infty < m \leq n\}$  is given by

$$f_D(z) = B(z)G(z) \quad (62)$$

$$= \frac{1}{\sigma_0 Q(z)} \left[ \frac{z^{-D}H^*(1/z^*)P_s(z)}{\sigma_0 Q^*(1/z^*)} \right]_+. \quad (63)$$

Since

$$r_{sx}[k] = E\left\{s[n] \left( \sum_{n=0}^{N-1} h[n]s[m-k-n] \right)^* \right. \\ \left. + E\{s[n]w^*[n-k]\} \right\} \quad (64)$$

$$= \sum_{n=0}^{N-1} h^*[n]r_s[k+n], \quad (65)$$

the resulting MSE is

$$\mathcal{J}_{\text{IIR,min}}^{(D)} = E\{(s[n-D] - \hat{s}[n-D])s^*[n-D]\} \quad (66)$$

$$= r_s[0] - E\left\{ \sum_{l=0}^{\infty} f_D[l]x[n-l]s^*[n-D] \right\} \quad (67)$$

$$= r_s[0] - \sum_{l=0}^{\infty} f_D[l]r_{sx}^*[l-D] \quad (68)$$

$$= r_s[0] - \sum_{l=0}^{\infty} \sum_{n=0}^{N-1} f_D[l]h[n]r_s^*[l-D+n] \quad (69)$$

$$= r_s[0] - \sum_{l=0}^{\infty} \sum_{n=0}^{N-1} f_D[l]h[n]r_s[D-l-n]. \quad (70)$$

## REFERENCES

- [1] H. S. Malvar, *Signal Processing with Lapped Transforms*. Boston, MA: Artech House, 1992.
- [2] P. P. Vaidyanathan, *Multirate Systems and Filter Banks*. Englewood Cliffs, NJ: Prentice-Hall, 1992.

# Fast Approximated Sub-Space Algorithms

Mohammed A. Hasan<sup>†</sup> and Ali A. Hasan<sup>‡</sup>

<sup>†</sup> Dept of Electrical & Computer Engineering, University of Minnesota Duluth

<sup>‡</sup>College of Electronic Engineering, Bani Waleed , Libya

## Abstract

*In this paper, fast techniques for invariant subspace separation with applications to the DOA and the harmonic retrieval problems are presented. The main feature of these techniques is that they are computationally efficient as they can be implemented in parallel and can be transformed into matrix inverse-free algorithms. The basic operations used are the QR factorization and matrix multiplication. Specifically, two types of methods are developed. The first method uses Newton-like iteration and is quadratically convergent. The second method can be developed to have convergence of any prescribed order. Using these approximations, the minimum norm solution for the DOA and the harmonic retrieval problems for the projection of least squares weight onto the signal subspace of the data is obtained simply, without performing any SVD. Some of the developed methods are also examined on several test problems.*

## 1. Introduction

The estimation of projections onto selective set of invariant subspaces of data and covariance matrices is a common requirement in the development of high resolution methods. This situation arises in adaptive processing of sensor array data or sum of sinusoids where the estimation of the number of strong signals present in a given set of data and the projections onto signal and noise subspaces is essential. Subspace based methods for frequency estimation rely on a low rank system model that is obtained by organizing the observed data samples into vectors. MUSIC and ESPRIT based estimators are then obtained using this vector model.

Projection of the least-squares weight vector onto subspace of reduced dimension is an established technique for reducing the number of adaptive degrees of freedom used by an adaptive sensor array. The main problem in conventional algorithms for subspace estimation based upon eigenvalue decomposition (EVD) or singular value decomposition (SVD) are, however, both expensive to compute and difficult to make recursive or implement in parallel. In contrast, algorithms based on the QR factorization have established pipelinable architectures.

Since many signal processing applications (e.g. projection beamforming, MUSIC) do not explicitly utilize the full set of signal eigenvalues, diagonalizing the co-

variance matrix of the data is not necessarily advantageous and is not required. Various alternatives were proposed by several authors. Kay and Shaw [1] suggested the use of polynomials and rational functions of the sample covariance matrix for approximating the signal subspace. In [2], Tufts and Melissinos used Lanczos and power-type methods to approximate the signal subspace. Karhunen and Joutsenalo [3] approximated the signal subspace using the discrete Fourier and Cosine transforms. Ermolaev and Gershman [4] used powers of sample covariance matrix based on Krylov subspaces to approximate the noise subspace when the number of impinging signals and a threshold which separates the signal and noise eigenvalues are known *a priori*. In this work, we assume that a rough estimate of a threshold is known. For useful articles and books, the reader is referred to [5], [6]-[8] and the references therein.

The proposed algorithms could prove useful if a threshold that separates noise and signal eigenvalues is known. This threshold can, in some cases, be obtained by tracking subspaces where largest eigenvalue of current noise subspace or smallest eigenvalues of current signal subspace of the power level of the noise floor are known. In these cases the proposed algorithm can help speed up the computation for final estimation of subspaces. Another application is when the rank of signal subspace is known.

## 2. Data Model

The  $N$  samples of a scalar valued signal  $y(n)$  are assumed to be the sum of  $M$  complex-valued sinusoids in additive zero mean white Gaussian noise

$$x_k(n) = \alpha_k e^{j(w_k n + \phi_k)}, \quad k = 1, 2, \dots, M,$$
$$y(n) = \sum_{k=1}^M x_k(n) + v(n), \quad n = 1, 2, \dots, N, \quad (1)$$

Here  $\alpha_k > 0$  is the amplitude and the frequencies  $w_1, \dots, w_M$  are assumed to be distinct parameters, and the phases  $\phi_k$  are assumed to be uniformly distributed on  $[0, 2\pi]$  and are mutually independent. The noise,  $v(n)$  is assumed to be independent of the phases and to satisfy

$$E\{v(n)v^*(n-k)\} = \sigma_v^2 \delta(k), \quad (2)$$

where  $(.)^*$  denotes complex conjugate and  $\delta(.)$  is the Kronecker delta function. A low rank matrix representation of the problem is obtained by collecting  $L > M$

received samples in a column vector

$$\mathbf{y}(n) = [y(n) \ y(n+1) \ \cdots \ y(n+L-1)]^T, \quad (3a)$$

where  $(\cdot)^T$  denotes the matrix transpose.

The notation  $\mathbf{x}(n)$  will denote the vector

$$\mathbf{x}(n) = [x_1(n) \ x_2(n+1) \ \cdots \ x_M(n+L-1)]^T. \quad (3b)$$

Hence  $\mathbf{y}(n)$  can be written as

$$\mathbf{y}(n) = \mathbf{V}(w)\mathbf{x}(n) + \mathbf{v}(n), \quad n = 1, \dots, N-L+1, \quad (4)$$

where the additive noise vector,  $\mathbf{v}(n)$ , is defined similarly to  $\mathbf{y}(n)$  in (3) and  $\mathbf{V}(w)$  is an  $L \times M$  Vandermonde matrix given by

$$\mathbf{V}(w) = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ e^{jw_1} & e^{jw_2} & \cdots & e^{jw_M} \\ \vdots & \vdots & \ddots & \vdots \\ e^{j(L-1)w_1} & e^{j(L-1)w_2} & \cdots & e^{j(L-1)w_M} \end{bmatrix}. \quad (5)$$

The argument  $w$  is omitted in the sequel when not required. The covariance matrix,  $R$ , of the received windowed sequence is

$$R_y = E\{\mathbf{y}(n)\mathbf{y}^*(n)\} = \mathbf{V}\mathbf{D}\mathbf{V}^* + \sigma_v^2 I_L, \quad (6)$$

where the covariance matrix  $\mathbf{D} = \text{diag}(\alpha_1^2, \dots, \alpha_M^2)$  is diagonal. The matrix  $I_L$  is the identity matrix of size  $L$ . Note that

$$R_x = E\{\mathbf{x}(n)\mathbf{x}^*(n)\} = \mathbf{V}\mathbf{D}\mathbf{V}^*. \quad (7)$$

Similar formulation can also be obtained for the direction of arrival (DOA) problem except in that case the matrix  $D$  is not necessarily diagonal.

In this paper, it is shown that if a threshold that separates the signal and noise eigenvalues or if the dimension of the signal subspace is a priori known, the subspace estimation can be obtained using the QR factorization of a large power of the covariance matrix.

### 3. Invariant Subspace Computation

Let  $A$  be a Hermitian matrix, and let  $P_0$  and  $P_1$  to denote the orthogonal projections onto the invariant subspaces corresponding to eigenvalues inside and outside the interval  $(-|b|, |b|)$ , where  $b$  is a nonzero. An elegant method for computing those invariant subspaces is presented next. Consider the sequence of matrices defined by

$$S_k = (b^k I_L - A^k)(b^k I_L + A^k)^{-1}, \quad (8)$$

then the eigenvalues of  $S_k$  given by  $\{\frac{b^k - \lambda_j^k}{b^k + \lambda_j^k}\}_{j=1}^L$  converge to 1 or -1 as  $k \rightarrow \infty$ . Thus  $S_k$  is bounded for all sufficiently large  $k$ . It can be shown the sequence  $S_k$  converges to a matrix  $S$  satisfying  $S^2 = I_L$ , and  $SA = AS$ . Moreover,  $S$  and  $A$  have the same invariant subspaces inside and outside a circle of radius  $|b|$  and centered at the origin. If (8) is computed directly using powers of the matrix  $A$ , over- and under-flow will occur.

Since the sample covariance matrix is generally positive semidefinite, we will apply this iteration on the shifted matrix. Fast implementation of computing the limit of the sequence  $\{S_k\}_{k=1}^\infty$  which also avoids the problem of over- and under-flow will be given next.

#### Algorithm 1:

$$S_0 = R_y - bI_L$$

$$S_{k+1} =$$

$$\{(I_L + S_k)^r - (I_L - S_k)^r\} \{(I_L + S_k)^r + (I_L - S_k)^r\}^{-1}. \quad (9)$$

It can be shown that  $S_k$  satisfies the following elegant error formula

$$\begin{aligned} (S_{k+1} + S)^{-1}(S_{k+1} - S) &= \{(S_k + S)^{-1}(S_k - S)\}^r \\ &= \{(S_0 + S)^{-1}(S_0 - S)\}^{r^k}. \end{aligned} \quad (10)$$

This method can be made to converge at any desired rate by choosing an appropriate  $r$ . From several numerical experiments, it was observed that for  $r = 2$ , a suitable  $K = 5$ , while  $K = 3$  if  $r = 3$ . Once the desired convergence is obtained, the signal subspace projection is computed as  $P_s = \frac{I_L + S_{K+1}}{2}$  and the noise subspace projection is approximated as  $P_n = \frac{I_L - S_{K+1}}{2}$ .

The next results provide quadratically convergent methods for subspace computation. The significance of the next theorem is that it computes the projection matrix for the subspaces whose eigenvalues fall between two numbers  $a$  and  $b$ .

**Theorem 1.** Let  $X_0 = R_y$  be a  $L \times L$  nonsingular matrix and let  $0 < a < b$  be two positive numbers. Let  $X_k$  be generated using

$$X_{k+1} = (2X_k - (a+b)I_L)^{-1}(X_k^2 - (a+b)X_k + abI_L), \quad (11a)$$

where  $I_L$  is the  $p \times p$  identity matrix. Then  $X_k$  converges quadratically to  $S = aQ_1 + bQ_2$ , where  $Q_1$  and  $Q_2$  are the projections onto the span of all eigenvectors of  $R_y$  whose corresponding eigenvalues are in the right and left half planes of the line which perpendicularly bisects the segment between  $a$  and  $b$ . Moreover,  $Q_1 = \frac{bI_L - S}{b-a}$ ,  $Q_2 = \frac{aI_L - S}{a-b}$  and  $X_k$  satisfies the following error formula

$$(X_{k+1} + S)^{-1}(X_{k+1} - S) = (X_k + S)^{-1}(X_k - S)^2. \quad (11b)$$

It should be stated that the above result holds true for any two numbers  $a \neq b$ . In this case if  $a + b = 0$  with  $a \neq 0$ , then the subspace decomposition reduces to computing the projections onto the subspaces spanned by the eigenvectors with eigenvalues having positive and negative real parts, respectively. Specifically, if  $a = -b = 1$ , the matrix  $S$  reduces to the matrix sign function of  $X_0$ .

When a threshold  $b$  which separates the signal and noise eigenvalues is a priori known, then the suggested approach will be very effective in extracting the signal and noise subspaces. More generally, one can derived a

stable and quadratically convergent algorithm for computing the invariant subspace of the matrix  $A$  in the half-plane with boundary determined by the line which perpendicularly bisects the line segment between  $z = 0$  and  $z = 2b$ .

**Theorem 2.** Let  $A$  be a nonsingular matrix of size  $L$  and let  $b \neq 0$  be a complex number. For  $k = 1, 2, \dots$ , compute

$$Z_{k+1} = \frac{1}{2} Z_k (Z_k - bI_L)^{-1} Z_k, \quad (12a)$$

with  $Z_1 = A$ . Then the sequence  $Z_k$  converges to  $2bZ$  where  $Z$  is the projection onto the subspace spanned by all eigenvectors whose eigenvalues are in the right half plane with boundary determined by the line which perpendicularly bisects the line segment between  $z = 0$  and  $z = 2b$ .

The quadratic convergence of this algorithm can be seen from the error formula which can be shown to be

$$(Z_{k+1} - 2bZ)Z_{k+1}^{-1} = \{(Z_k - 2bZ)Z_k^{-1}\}^2. \quad (12b)$$

Note that the matrix inverse in (2a) can be avoided by utilizing the Schultz iteration [9].

The main disadvantage of (9) and (12) is that they require the computation of matrix inverse. In the following result an implementation of (9) which avoids matrix inverse computation is given.

**Theorem 3.** Let  $b$  be a threshold which separate the signal and noise eigenvalues of the positive definite matrix  $R_y$ . Let  $S_k$  be a sequence generated as follows:

$$\begin{aligned} S_0 &= R_y - bI_L \\ \begin{bmatrix} (I_L + S_k)^r \\ (I_L - S_k)^r \end{bmatrix} &= \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix} \begin{bmatrix} R_k \\ 0 \end{bmatrix}, \quad k = 0, 1, 2, \dots, K \\ S_{k+1} &= (Q_{11} + Q_{21})(Q_{11}^* - Q_{21}^*), \end{aligned} \quad (13)$$

then  $S_k$  converges to  $P_{\lambda < |b|} - P_{\lambda > |b|}$ .

Note that the middle step in Equation (13) involves QR decomposition. This provides an  $r$ th order convergent algorithm for computing the projections onto invariant subspaces to the left and right of the line  $z = b$ . Once  $S$  is computed accurately, then the eigen-spaces can be obtained from the QR factorization of  $\frac{I_L + S}{2}$ , i.e.,  $\frac{I_L + S}{2} = QR$ , then  $Q^*(R_y - bI_L)Q = \begin{bmatrix} A_1 & 0 \\ 0 & A_2 \end{bmatrix}$ , where all eigenvalues of  $A_1$  are inside the interval  $[-|b|, |b|]$  and those of  $A_2$  are outside that interval. This process can be repeated if necessary on smaller matrices  $A_1$  and  $A_2$ . Initial tests of this algorithm have shown that this implementation is stable and convergent even when the matrix  $A$  has an eigenvalue as small in magnitude as  $10^{-13}$ .

We should note that in Iteration (13), orthogonal projections are obtained using only matrix multiplication and the QR factorization. This method can be made to converge at any desired rate by choosing an appropriate  $r$ .

## Algorithm 2:

Using analogous derivation, we obtained another inverse-free implementation of (13) for Hermitian matrices which is given as follows:

$$\begin{aligned} P_0 &= R_y - bI_L \\ \begin{bmatrix} P_k^r \\ (I_p - P_k)^r \end{bmatrix} &= \begin{bmatrix} Q_{11} & Q_{12} \\ Q_{21} & Q_{22} \end{bmatrix} \begin{bmatrix} R_k \\ 0 \end{bmatrix}, \quad k = 0, 1, 2, \dots, K \\ P_{k+1} &= Q_{11}(Q_{11}^* - Q_{21}^*), \end{aligned} \quad (14)$$

then  $P_k$  converges to an orthogonal projection. Let  $P_{K+1} = QR$  be a QR factorization, then  $Q^*AQ$  is block diagonal. This algorithm indicates that projections onto half-planes can be obtained using only matrix multiplication and QR factorization.

## 4. Estimation of a Threshold

The performance of estimators based on the approximations given in the previous section is mainly dependent on the accuracy of a threshold that separates the signal and noise eigenvalues or if the dimension of the signal subspace is a priori known.

Since  $R_y$  is Hermitian, it has the eigendecomposition  $R_y = \sum_{i=1}^L \lambda_i u_i u_i^*$  where  $\lambda_i$  and  $u_i$  are the  $i$ th eigenvalue and  $i$ th corresponding eigenvector. For convenience, it is assumed that the eigenvalues are sorted in decreasing order so that  $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_M > \lambda_{M+1} = \dots = \lambda_L = \sigma_v^2$  with corresponding eigenvectors  $\{u_i\}_{i=1}^L$ . The eigenvectors  $\{u_i\}_{i=1}^M$  are usually called the signal vectors and the eigenvectors  $\{u_i\}_{i=M+1}^L$  are called the noise vectors. If the average of the signal eigenvalues is denoted by  $\bar{\lambda}_s$ , then one can show that  $\frac{\text{trace}(R_y)}{L}$  is a good estimate of the threshold provided that  $L$  is sufficiently large. The main requirement for this threshold is  $\sigma_v^2 < \frac{\text{trace}(R_y)}{L} < \lambda_M$  which holds provided

$$\frac{L-M}{M}(\lambda_M - \sigma_v^2) > \bar{\lambda}_s - \lambda_M. \quad (15)$$

Note that in this inequality the only parameter that can be varied is  $L$ . Clearly, if  $L$  is much larger than  $M$  so that  $\frac{L-M}{M} \gg 1$ , then the above inequality will hold. Although this threshold is very simple to compute, it holds only for the theoretical covariance matrix, i.e., all noise eigenvalues are the same. Another observation is that (15) holds for smaller  $L$  if the spread of signal eigenvalues is small and thus the difference  $\bar{\lambda}_s - \lambda_M$  is small, or if  $\lambda_M - \sigma_v^2$  is large. Both of these cases lead to smaller  $L$  for (15) to hold.

Note that for  $M = 2$ , (15) reduces to

$$\frac{L-2}{2}(\lambda_2 - \sigma_v^2) > \lambda_1 - \lambda_2.$$

Also in the hypothetical case in which all signal eigenvalues are equal the above threshold always accurate for any  $L > M$ .

When  $\bar{\lambda}_s - \lambda_M$  is large, one can use a sharper estimate of the threshold based on

$$\mu = \frac{\sum_{i=1}^L \sqrt{\lambda_i}}{L} = \frac{T}{L}.$$

This estimate can be computed from the covariance matrix but the computation is very lengthy and complicated even when  $L$  is low. For example, when  $L = 2$  the value of  $\mu$  can be estimated from

$$T^2 = \text{trace}(R_y) + 2\sqrt{\det(R_y)},$$

where  $T = \mu L$ . For  $L = 3$ ,  $T$  can be estimated by solving the following equation

$$\{(T^2 - a)^2 - 4b\}^2 = 8\sqrt{c}T,$$

where  $a, b, c$  are determined from the characteristic polynomial of  $R_y$  given by  $\lambda^3 - a\lambda^2 + b\lambda - c$ .

## 5. Simulation Results

In this section, frequency estimators based on subspace approximations are examined on several data sets generated by the equation

$$y(n) = d_1 e^{j(2\pi f_1 n + \phi_1)} + d_2 e^{j(2\pi f_2 n + \phi_2)} + v(n), \quad (15)$$

where  $d_1 = 1.0$ ,  $d_2 = 1.0$ ,  $f_1 = 0.5$ ,  $f_2 = 0.52$  and  $n = 1, 2, \dots, N = 25$ . The  $\phi_i$  are independent random variables uniformly distributed over the interval  $[-\pi, \pi]$ . The noise  $v(k)$  is assumed to be white and uncorrelated with the signal. Note that  $f_2 - f_1 < \frac{1}{N}$ .

The SNR for either sinusoids is defined as  $10 \log_{10}(\frac{\sigma_x^2}{\sigma_v^2})$ , where  $x(n) = d_1 e^{j(2\pi f_1 n + \phi_1)} + d_2 e^{j(2\pi f_2 n + \phi_2)}$  and  $\sigma_x^2$ ,  $\sigma_v^2$  are the variances of  $x(n)$  and  $v(n)$ , respectively. The size of the covariance matrix is chosen to be  $L = 10$  which in the absence of noise has effective rank two. We performed experiments to compare the proposed methods versus the truncated SVD-based MUSIC. The SVD routine on MATLAB is used for the computation of the signal subspace eigenvectors and eigenvalues required to implement a SVD-based method for comparison. We varied SNR from 10 to 20 in 5dB steps and estimated the frequencies for data length 25. For each experiment (with data length and SNR fixed), we performed 100 independent trials to estimate the frequencies. We use the following performance criterion (RMSE)

$$RMSE = \sqrt{\frac{1}{N_e} \sum_{i=1}^{N_e} (\hat{f}_i - f_{true})^2}$$

to compare the results. Here  $N_e$  is the number of independent realizations, and  $\hat{f}_i$  is the estimate provided from the  $i$ th realization. Several experiments were conducted to test the performance of the algorithms presented in Theorem 3, and the SVD-based MUSIC. The mean values of estimated frequencies and their RMSE of the SVD-based MUSIC are given in Table 1.

SNR	$f_1$	$f_2$	$RMSE_{f_1}$	$RMSE_{f_2}$
20 dB	0.500556	0.522322	0.00563	0.012522
15 dB	0.500729	0.521735	0.00652	0.014531
10 dB	0.500961	0.524952	0.00813	0.019204

Table 1: Mean and RMSE of frequencies for data of two complex sinusoids at frequencies 0.50 and 0.52 in noise with SNR=20, 15, 10 dB, dimension of data vectors  $L=10$ . Theorem 3 is used.

## References

- [1] Kay S. M. and Shaw A. K., "Frequency Estimation by Principal Component AR Spectral Estimation Method without Eigendecomposition," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, Vol. 36, No. 1, pp. 95-101, January 1988.
- [2] Tufts D. and Melissinos C. D., "Simple, Effective Computation of Principal Eigenvectors and Their Eigenvalues and Application to High-Resolution Estimation of frequencies," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, Vol. 34, No. 5, pp. 1046-1053, October 1986.
- [3] Karhunen J. T., and Joutsenalo J., "Sinusoidal Frequency Estimation by signal subspace Approximation," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, Vol. ASSP-40, No. 12, pp. 2961-2972, December 1992.
- [4] Ermolaev V. T. and Gershman A. B., "Fast Algorithm for Minimum-Norm Direction-of-Arrival Estimation," *IEEE Trans. on Signal Processing*, Vol. 42, No. 9, pp. 2389-2394, September 1994.
- [5] Kay S. M., *Modern Spectral Estimation, Theory and Applications*, Englewood Cliffs, NJ: Prentice-Hall, 1988.
- [6] Hasan M. A., and Hasan A. A., "Hankel Matrices of Finite Rank with Applications to Signal Processing and Polynomials", *J. of Math. Anal. and Appls*, 208, pp.218-242, 1997.
- [7] Hasan M. A., Azimi-Sadjadi M. R., "Separation of Multiple Time Delays Using New Spectral Estimation Schemes with Applications to Underwater Target Detection", *IEEE Trans. on Signal Processing*, Vol. 46, No. 6, pp. 1580-1590, June 1998.
- [8] Hasan M. A., "DOA and Frequency Estimation Using Fast Sub-Space Algorithms," accepted for publication in *Journal of Signal Processing*.
- [9] Stoer J. and Bulirsch, *Introduction to Numerical Analysis*, Springer-Verlag, New York 1980.

# STOCHASTIC ALGORITHMS FOR MARGINAL MAP RETRIEVAL OF SINUSOIDS IN NON-GAUSSIAN NOISE

Christophe Andrieu - Arnaud Doucet

Signal Processing Group, University of Cambridge  
Department of Engineering, Trumpington Street  
CB2 1PZ Cambridge, UK

Email: ca226@eng.cam.ac.uk - ad2@eng.cam.ac.uk

## ABSTRACT

In this paper we propose a method to estimate the frequencies of sinusoids embedded in non-Gaussian noise. We model the noise using mixtures of Gaussians and propose two original, efficient algorithms that allow for the marginal MAP estimation of the sinusoid parameters to be estimated. Outline of the proof of convergence of the algorithms is also given and simulation results are presented.

## 1 Introduction

The harmonic retrieval problem is a fundamental problem in signal processing that has numerous applications in radar, seismology and nuclear magnetic resonance. Many efforts have been devoted to the development of methods that address this problem, ranging from periodogram related procedures, to subspace and parametric methods relying on maximum likelihood or Bayesian estimation. The Bayesian estimation of harmonic signals in white Gaussian noise has been the subject of many recent papers, see [1], [2], [4], [5], among others. Here we address the important and more difficult problem of estimating the frequencies of sinusoids embedded in non-Gaussian noise, and formulate it in a Bayesian framework. A commonly used tool to model non-Gaussian distributions consists of using discrete or continuous mixtures of Gaussian distributions, and this is the approach adopted here. The motivation for this choice is that by introducing a proper set of (artificial) missing data, say  $\xi$ , one can often design simple and efficient algorithms that allow for the estimation of important features of the posterior distribution related to the problem. However, from a statistical point of view the introduction of missing data can typically lead to inconsistent estimators as the number of parameters to be estimated typically grows with the number of observations. Joint estimators, *i.e.* estimators involving  $\xi$ , should thus be avoided and marginal estimation of the parameters should be favoured.

For the case of sinusoids embedded in a noise modelled as a mixture of Gaussians, the analytical expression of the marginal posterior distribution of interest is of the form

$$p(\mathbf{a}, \omega, \delta | \mathbf{y}) = \int p(\mathbf{a}, \omega, \delta, \xi | \mathbf{y}) d\xi,$$

where  $\mathbf{a}, \omega$  are the amplitude and pulses of the sinusoids and  $\delta$  are parameters of the observations noise. Unfortunately it is not

available in closed-form and one has to resort to numerical methods. Monte Carlo methods, and in particular Monte Carlo Markov chain methods (MCMC) have proved to be efficient tools for the estimation of certain features of complicated posterior distributions, in particular MMSE (Minimum Mean Square Error) *e.g.*  $\mathbb{E}((\mathbf{a}, \omega, \delta) | \mathbf{y})$  in the case treated here.

However this choice of estimator is not adapted when the marginal posterior distribution is multimodal and the MMSE estimate located between the modes, possibly in a region of very low probability. Computing MAP (Maximum *a posteriori*) estimates of the frequencies might be preferable in such cases, but whereas MCMC methods are well adapted to the estimation of marginal posterior means, their use to perform MMAP (Marginal MAP) estimation can be questionable. Indeed in this case further approximations are introduced by histogram or density estimation methods and require careful tuning of extra parameters.

The EM (Expectation Maximization) algorithm is designed to converge towards a stationary point of the marginal posterior distribution. It is however limited to certain classes of models for which the expectation and maximization steps can be performed conveniently. This is why stochastic versions have been proposed, such as SEM (Stochastic EM) or MCEM (Monte Carlo EM). Convergence results are sparse and the algorithms do not always fully exploit the structure of the statistical model. In this paper we propose several Monte Carlo methods for performing MMAP of the frequencies of sinusoids embedded in non-Gaussian noise. The first method relies on the SAME (State Augmentation for Marginal Estimation) algorithm [10]. This algorithm is conceptually very simple and straightforward to implement in most cases, requiring only small modifications to MCMC code written for sampling from  $p(\mathbf{a}, \omega, \delta, \xi | \mathbf{y})$ . In order to reduce the computational complexity of this algorithm, we present a stochastic approximation type extension of this algorithm. We then present an original analysis of the convergence of the stochastic approximation type algorithm which relies on a perturbation analysis of the original SAME algorithm. Simulation results are presented that demonstrate the interest of the approach.

This paper is organized as follows. In Section 2 the signal models are given. In Section 3, we formalize the Bayesian model and specify the prior distributions. Section 4 is devoted to Bayesian computation. We propose non homogeneous MCMC algorithms to perform Bayesian inference for which sufficient condition for global convergence can be established. Performance of these algorithms is illustrated by computer simulations on synthetic data in

C. Andrieu is sponsored by AT&T Laboratories, Cambridge UK. A. Doucet is sponsored by EPSRC, UK.

## 2 Problem statement

Let  $\mathbf{y} \triangleq (y_1, y_2, \dots, y_T)^T$  be an observed vector of  $T$  real data samples. The elements of  $\mathbf{y}$  are the superimposition of  $k$  sinusoids corrupted by noise  $\mathbf{n} \triangleq (n_1, \dots, n_T)^T$ :

$$y_t = \sum_{j=1}^k a_{c_j} \cos(\omega_j t) + a_{s_j} \sin(\omega_j t) + n_t,$$

where  $1 \leq k \leq \lfloor (T-1)/2 \rfloor$ ,  $a_{c_j}$ ,  $a_{s_j}$  and  $\omega_j$  are respectively the amplitudes and the radial frequency of the  $j^{\text{th}}$  sinusoid. We assume that  $\omega \in \Omega \triangleq \{\omega \in (0, \pi)^k; \omega_{j_1} \neq \omega_{j_2} \text{ for } j_1 \neq j_2\}$ . In a vector-matrix form, we have

$$\mathbf{y} = \mathbf{D}(\omega) \mathbf{a} + \mathbf{n},$$

where  $[\mathbf{a}]_{2i-1,1} \triangleq a_{c_i}$ ,  $[\mathbf{a}]_{2i,1} \triangleq a_{s_i}$  and  $[\omega]_{i,1} \triangleq \omega_i$  for  $i = 1, \dots, k$ . The  $T \times 2k$  matrix  $\mathbf{D}(\omega)$  is defined as  $[\mathbf{D}(\omega)]_{t,2j-1} = \cos[\omega_j t]$  and  $[\mathbf{D}(\omega)]_{t,2j} = \sin[\omega_j t]$  for  $t = 1, \dots, T$ , and  $j = 1, \dots, k$ . The noise is assumed white, distributed according to a mixture of Gaussian distributions, *i.e.*<sup>1</sup>

$$n_t \stackrel{iid}{\sim} \lambda \mathcal{N}(0, \sigma^2) + (1 - \lambda) \mathcal{N}(0, \alpha \sigma^2),$$

where  $0 < \lambda < 1$  defines the mixture probability,  $\sigma^2$  is a global scale parameter and  $0 < \alpha < 1$ . It is convenient to introduce the so-called missing data  $\mathbf{r}_{1:T}$  such that:

$$n_t | r_t \sim \mathcal{N}(0, \sigma^2 \mathbb{I}_{\{1\}}(r_t) + \alpha \sigma^2 \mathbb{I}_{\{0\}}(r_t)),$$

and  $\Pr(r_t = 1) = \lambda$  and  $\Pr(r_t = 0) = 1 - \lambda$ . This allows us to write the likelihood of the observations

$$p(\mathbf{y} | \mathbf{a}, \omega, \lambda, \mathbf{r}_{1:T}, \alpha, \sigma^2) = |2\pi\sigma^2\boldsymbol{\Sigma}|^{-1/2} \times \exp\left(-\frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{D}(\omega) \mathbf{a})^T \boldsymbol{\Sigma}^{-1} (\mathbf{y} - \mathbf{D}(\omega) \mathbf{a})\right),$$

where  $\boldsymbol{\Sigma} \triangleq \text{diag}(\mathbb{I}_{\{1\}}(r_j) + \alpha \mathbb{I}_{\{0\}}(r_j))$   $j = 1, \dots, T$ . Note that this likelihood is invariant by permutation of the indexes of the pulses  $\omega_j$ , if no ordering constraint is introduced, and that consequently MMSE estimates can lead to very poor results. The parameters of the sinusoids, of the noise and the missing data *i.e.*  $\boldsymbol{\theta} \triangleq (\mathbf{a}, \omega, \lambda, \mathbf{r}_{1:T}, \alpha, \sigma^2)$  are unknown, and our aim is to estimate these parameters;  $\mathbf{a}$  and  $\omega$  being in general the parameters of primary interest. Note that the strategy developed in this paper can be extended to the case of continuous Gaussian mixtures, in order to model heavy tailed distributions, but we do not consider this case here.

## 3 Bayesian Models and Estimation Objectives

In this paper we follow a Bayesian approach where the unknown parameter vector  $\boldsymbol{\theta}$  is regarded as being drawn from an appropriate prior distribution. This prior distribution reflects our degree of belief in the relevant values of the parameters. Note that when no prior knowledge is available, then uninformative distributions can be used [3]. This is the approach we follow here. We first propose a model that sets up a probability distribution over the space of possible structures of the signal and we give the estimation aims.

<sup>1</sup>This could be extended to the case of discrete mixtures with more components.

## 3.1 Prior distribution

We set a prior distribution on the unknown parameter vector  $\boldsymbol{\theta} = (\mathbf{a}, \omega, \lambda, \mathbf{r}_{1:T}, \alpha, \sigma^2) \in \Theta$  where  $\Theta \triangleq \mathbb{R}^{2k} \times \Omega \times (0, 1) \times \{0, 1\}^T \times (0, 1) \times \mathbb{R}^+$ . The following uninformative improper prior distribution<sup>2</sup> is selected:

$$p(\mathbf{a}, \omega, \sigma^2 | \mathbf{r}_{1:T}, \alpha) \propto \left| \frac{\mathbf{D}^T(\omega) \boldsymbol{\Sigma}^{-1} \mathbf{D}(\omega)}{\sigma^2} \right|^{1/2} \mathbb{I}_{\Omega}(\omega).$$

This prior corresponds to Jeffreys' prior for the linear model [3]. It penalizes close frequencies as pointed out in [5]. The parameters  $\alpha$  and  $\lambda$  are assumed distributed according to  $\alpha \sim \mathcal{U}_{(0,1)}$  and  $\lambda \sim \mathcal{U}_{(0,1)}$  which are vague prior distributions.

## 3.2 Estimation objectives

Given the observations  $\mathbf{y}$ , Bayesian inference about  $\boldsymbol{\theta}$  is based on the posterior distribution  $p(\boldsymbol{\theta} | \mathbf{y})$  obtained from Bayes' theorem,

$$p(\boldsymbol{\theta} | \mathbf{y}) \propto p(\mathbf{y} | \boldsymbol{\theta}) p(\boldsymbol{\theta}).$$

Our aim is to estimate this joint distribution from which, by standard probability marginalization and transformation techniques, one can "theoretically" obtain all posterior features of interest including the marginal distributions, posterior modes or conditional expectations such as the MMSE estimate

$$\mathbb{E}[\boldsymbol{\theta} | \mathbf{y}] = \int_{\Theta} \boldsymbol{\theta} p(\boldsymbol{\theta} | \mathbf{y}) d\boldsymbol{\theta},$$

among others. As discussed in the introduction this problem can be addressed using MCMC methods but the use of these techniques for the computation of the MMAP estimator  $(\mathbf{a}, \omega, \sigma^2, \alpha)_{\text{MMAP}}$  defined as

$$\arg \max_{(\mathbf{a}, \omega, \sigma^2, \alpha) \in \mathbb{R}^{2k} \times \Omega \times \mathbb{R}^+ \times (0, 1)} p(\mathbf{a}, \omega, \sigma^2, \alpha | \mathbf{y}),$$

can be questionable. In the next section we describe an algorithm that allows for computation to be performed by adapting MCMC techniques for MMAP estimation.

## 4 Bayesian Marginal MAP robust spectral estimation

### 4.1 The SAME algorithm

One might be interested in the marginal MAP estimation of the frequencies, *i.e.* finding the maximum of  $p(\mathbf{a}, \omega, \sigma^2, \alpha | \mathbf{y})$ . In order to achieve this we introduce two versions of the SAME algorithm [10], the second one being a stochastic approximation type algorithm. Let us consider the extended probabilistic model,

$$\begin{aligned} & \bar{p}^{\otimes \gamma}(\mathbf{a}, \omega, \sigma^2, \alpha, \lambda_{1:\gamma}, \mathbf{r}_{1:T,1:\gamma} | \mathbf{y}) \\ & \propto \prod_{j=1}^{\gamma} p(\mathbf{y} | \mathbf{a}, \omega, \sigma^2, \alpha, \lambda_j, \mathbf{r}_{1:T,j}) p(\mathbf{a}, \omega, \sigma^2, \alpha, \lambda_j, \mathbf{r}_{1:T,j}), \end{aligned}$$

where  $\gamma$  is a positive integer,  $\mathbf{r}_{1:T,j}$  is a replica of the missing data. Clearly this probabilistic model admits  $\bar{p}^{\gamma}(\mathbf{a}, \omega, \sigma^2, \alpha | \mathbf{y})$  as marginal distribution, where  $\bar{p}^{\gamma}(\mathbf{a}, \omega, \sigma^2, \alpha | \mathbf{y})$  is the distribution proportional to  $p^{\gamma}(\mathbf{a}, \omega, \sigma^2, \alpha | \mathbf{y})$ . Given a sequence  $(\gamma_i)_{i \in \mathbb{N}}$  such that  $\lim_{i \rightarrow +\infty} \gamma_i = +\infty$ , the idea of the SAME algorithm is to run a non homogeneous Markov chain that admits  $\bar{p}^{\otimes \gamma_i}(\mathbf{a}, \omega, \sigma^2, \alpha | \mathbf{y})$  as invariant distribution at each iteration  $i$ .

<sup>2</sup>A prior distribution  $p(\boldsymbol{\theta})$  is said to be improper if  $\int_{\Theta} p(\boldsymbol{\theta}) d\boldsymbol{\theta} = +\infty$ .



The distribution  $\bar{p}^{\otimes \gamma_i}(\mathbf{a}, \omega, \sigma^2, \alpha | \mathbf{y})$  concentrates itself on its set of global maxima as  $i \rightarrow +\infty$  (this is the idea of simulated annealing) and the algorithm is thus hoped in practice to converge towards a global maximum. Note that when  $\gamma_i = 1$  for  $i \geq 1$  this algorithm is a standard MCMC that asymptotically produces samples from  $p(\theta | \mathbf{y})$ . In practice one can make use of the properties of the model and analytically integrate out  $\mathbf{a}, \sigma^2$  and  $\lambda_i$ , leading to an expression of  $\bar{p}^{\otimes \gamma}(\omega, \alpha, \mathbf{r}_{1:T, 1:\gamma} | \mathbf{y})$  up to a constant. It can be shown that

$$\begin{aligned} \bar{p}^{\otimes \gamma}(\omega, \alpha, \mathbf{r}_{1:T, 1:\gamma} | \mathbf{y}) &\propto \prod_{j=1}^{\gamma} |\mathbf{D}^T \Sigma_j^{-1} \mathbf{D}|^{1/2} |\Sigma_j|^{-1/2} \\ &\times |\mathbf{M}_\gamma|^{-1/2} [\mathbf{y}^T \mathbf{P}_\gamma(\omega) \mathbf{y}]^{-\gamma(T/2+k)+k} \\ &\times \prod_{j=1}^{\gamma} n_{1,j}! (T - n_{1,j})!, \end{aligned}$$

where  $n_{1,j} \triangleq \sum_{t=1}^T \mathbb{I}_{\{0\}}(r_{t,j})$  and

$$\begin{aligned} \mathbf{M}_\gamma &= [\mathbf{D}^T(\omega) \Psi_\gamma^{-1} \mathbf{D}(\omega)]^{-1}, \\ \mathbf{m}_\gamma &= \mathbf{M}_\gamma \mathbf{D}^T(\omega) \Psi_\gamma^{-1} \mathbf{y}, \quad \Psi_\gamma^{-1} = \sum_{i=1}^{\gamma} \Sigma_i^{-1}, \\ \mathbf{P}_\gamma(\omega) &= \Psi_\gamma^{-1} - \Psi_\gamma^{-1} \mathbf{D}(\omega) \mathbf{M}_\gamma \mathbf{D}^T(\omega) \Psi_\gamma^{-1}. \end{aligned}$$

In order to sample from  $\bar{p}^{\otimes \gamma}(\omega, \alpha, \mathbf{r}_{1:T, 1:\gamma} | \mathbf{y})$ , we propose the following algorithm:

#### MCMC algorithm for marginal spectral analysis

1. Initialization  $\theta^{(0)} = \{\omega^{(0)}, \alpha^{(0)}, \mathbf{r}_{1:T, 1:\gamma_0}^{(0)}\}$  and  $i = 1$ .
2. Iteration  $i$ 
  - For  $j = 1, \dots, \gamma_i$ , sample  $r_{t,j}^{(i)}$  from  $\bar{p}^{\otimes \gamma_i}(r_{t,j} | \mathbf{y}, \omega^{(i-1)}, \alpha, \mathbf{r}_{1:T, 1:\gamma_0}^{(i-1)}, \mathbf{a}^{(i-1)}, \sigma^{2(i-1)})$  for  $t = 1, \dots, T$ .
  - Sample  $\alpha^{(i)} \sim \bar{p}^{\otimes \gamma_i}(\alpha | \mathbf{y}, \omega^{(i-1)}, \mathbf{r}_{1:T}^{(i)}, \mathbf{a}^{(i-1)}, \sigma^{2(i-1)})$ .
  - Sample  $\omega_j^{(i)} \sim \bar{p}^{\otimes \gamma_i}(\omega_j | \mathbf{y}, \omega_{\gamma_j}^{(i-1)}, \alpha^{(i)}, \mathbf{r}_{1:T, 1:\gamma_i}^{(i)})$  for  $j = 1, \dots, k$  with an MCMC step.
  - Sample  $\mathbf{a}^{(i)}, \sigma^{2(i)} \sim \bar{p}^{\otimes \gamma_i}(\mathbf{a}, \sigma^2 | \mathbf{y}, \omega^{(i)}, \alpha^{(i)}, \mathbf{r}_{1:T, 1:\gamma}^{(i)})$ .

Where  $\mathbf{r}_{-i}$  means “ $\mathbf{r}_{1:T}$  with  $r_i$  removed” and similarly for  $\omega_{-j}$ . We comment the different sampling steps:

- Sampling  $r_{t,j}$  is straightforward as it simply involves sampling from a discrete distribution.
- Sampling  $\omega_j$  can be done using an adaptation of the technique described in [1].
- Sampling  $\mathbf{a}, \sigma^2$  is standard as it requires the simulation from an inverse-Gamma distribution and a normal distribution.
- Sampling  $\alpha$  mainly amounts to sampling from a truncated inverse-Gamma distribution and can be done efficiently by using a rejection method based on the work of [8].

This elegant algorithm allows to sample from the series of distributions of interest and convergence results can be proved that support the validity of the approach (See Section 5). However we see that as  $\gamma_i$  approaches infinity the computational burden of the algorithm becomes rapidly unrealistic. Thus we propose here a stochastic approximation adaptation of the algorithm presented above, which is computationally much cheaper.

## 4.2 The SA<sup>2</sup>ME

In the current version of the SAME algorithm  $\gamma_i$  replicas of the variables  $\mathbf{r}_{1:T}$  are sampled at each iteration  $i$ , which can rapidly become cumbersome as  $\gamma_i$  becomes large. Let  $i_0$  be an iteration chosen by the user. Then we propose from iteration  $i_0$  not to resample the variables  $\mathbf{r}_{1:T, 1:\gamma_{i-1}}$  that are “frozen” once they are simulated but simply sample the new replicas  $\mathbf{r}_{1:T, \gamma_{i-1}+1:\gamma_i}$ . The computational gain of this SA<sup>2</sup>ME (Stochastic Approximation SAME) is obvious and the analogy with classical stochastic approximation algorithms is clear, although we here take advantage of the statistical structure of the problem. However the algorithm is no longer a Markov chain as the update of the parameters at iteration  $i$  depends on the past of the chain up to iteration  $i-1$ . In fact this new algorithm can be viewed as a perturbation of the original SAME algorithm, and an analysis of these perturbations can be carried out to prove the validity of the new scheme, as sketched in the next section.

## 5 Convergence analysis

We first point out a convergence result for the SAME algorithm and then focus on the SA<sup>2</sup>ME algorithm.

### 5.1 SAME algorithm

First we set  $\theta_1 = \{\mathbf{a}, \omega, \sigma^2, \alpha\}$  and  $\theta_2 = \{\lambda, \mathbf{r}_{1:T}\}$  and name their state spaces  $\Theta_1$  and  $\Theta_2$ . The SAME algorithm defines a Markov chain on  $\theta_1$ , and it can be proved that this Markov chain is uniformly ergodic for a constant sequence  $\gamma_i$ , i.e. for any probability distribution  $\mu$ ,

$$\lim_{i \rightarrow +\infty} \|\mu K_{\gamma_i}^n(d\theta_1) - \bar{p}^{\gamma_i}(d\theta_1)\| = 0,$$

at a geometric rate independent of the initial condition, where  $\|\cdot\|$  is the total variation norm. Here  $K_i$  is the transition kernel of the SAME algorithm at iteration  $i$  which can be formally written as

$$K_i(\theta_1^{(i-1)}; \theta_1^{(i)}) \propto \int_{\Theta_2} p(\theta_1^{(i)} | \theta_2^{(1:\gamma_i)}) \prod_{j=1}^{\gamma_i} p(d\theta_2^{(j)} | \theta_1^{(i)}).$$

This convergence result mainly relies on the fact that the parameters  $\theta_1$  and  $\theta_2$  lie in bounded sets. From this result and following arguments similar to that used to prove the convergence of simulated annealing, it can be shown that for a logarithmic series of  $\gamma_i$  the SAME algorithm for MMAP estimation converges in the following sense

$$\lim_{n \rightarrow +\infty} \|\mu K_1 K_2 \dots K_n(d\theta_1) - \bar{p}^{\gamma_n}(d\theta_1)\| = 0.$$

Furthermore as the sequence  $\bar{p}^{\gamma_i}(d\theta_1)$  tends to a mixture of delta functions located at the global maxima of  $p(\theta_1)$  we conclude that the algorithm will asymptotically provide us with an estimate of  $\theta_{1, \text{MMAP}} \triangleq \arg \max_{\theta_1 \in \Theta_1} p(\theta_1)$ .

### 5.2 SA<sup>2</sup>ME

The proof of convergence of the algorithm relies on an analysis of the perturbations introduced by the new scheme upon the original SAME algorithm. We sketch here the proof of the algorithm, outline the main propositions that lead to the convergence result and explain their intuitive meaning. We introduce some notation that

will be useful throughout the proof. We introduce the transition probability corresponding to the SA<sup>2</sup>ME algorithm

$$\begin{aligned} & \tilde{K}_{i+1} \left( \tilde{\theta}_2^{(1:\gamma_i)}, \theta_1^{(i)}; d\tilde{\theta}_2^{(\gamma_i+1:\gamma_{i+1})}, d\theta_1^{(i+1)} \right) \\ & \propto p \left( d\theta_1^{(i+1)} \middle| \tilde{\theta}_2^{(1:\gamma_i)}, \tilde{\theta}_2^{(\gamma_i+1:\gamma_{i+1})} \right) p \left( d\tilde{\theta}_2^{(\gamma_i+1:\gamma_{i+1})} \middle| \theta_1^{(i)} \right), \end{aligned}$$

Here we simply express the fact that the missing data  $\tilde{\theta}_2^{(1:\gamma_i)}$  are “frozen” once they are simulated. In order to study the convergence properties of the second algorithm it will be useful to introduce for some integer  $k$  the transition kernel of the algorithm for which only the missing data up to iteration  $k-1$ ,  $\tilde{\theta}_2^{(1:\gamma_{k-1})}$ , are frozen, and missing data from then on,  $\tilde{\theta}_2^{(\gamma_{k-1}+1:\gamma_{i+1})}$ , are sampled at each iteration. More precisely, for  $i > k$  we define

$$\begin{aligned} & \hat{K}_{i+1,k} \left( \tilde{\theta}_2^{(1:\gamma_{k-1})}, \theta_1^{(i)}; d\theta_1^{(i+1)} \right) \\ & \propto \int_{\Theta_2^{\gamma_{i+1}-\gamma_{k-1}}} p \left( d\theta_1^{(i+1)} \middle| \tilde{\theta}_2^{(1:\gamma_{k-1})}, \theta_2^{(\gamma_{k-1}+1:\gamma_{i+1})} \right) \\ & \quad \times \prod_{j=\gamma_{k-1}+1}^{\gamma_{i+1}} p \left( d\theta_2^{(j)} \middle| \theta_1^{(i)} \right). \end{aligned}$$

In order to study the convergence properties of our algorithm, we will need notation to combine these kernels, namely,

$$\begin{aligned} & \mu \tilde{K}_{1:n} \left( d\theta_1^{(n)}, d\theta_2^{(\gamma_{n-1}+1:\gamma_n)} \right) = \int_{\Theta_1^n \times \Theta_2^{\gamma_n-1}} \mu \left( d\theta_1^{(0)} \right) \\ & \times \tilde{K}_1 \left( \theta_1^{(0)}; d\theta_1^{(1)}, d\tilde{\theta}_2^{(1)} \right) \tilde{K}_2 \left( \theta_1^{(1)}, \tilde{\theta}_2^{(1)}; d\theta_1^{(2)}, d\tilde{\theta}_2^{(2:\gamma_2)} \right) \\ & \dots \times \tilde{K}_j \left( \theta_1^{(j-1)}, \tilde{\theta}_2^{(1:\gamma_{j-1})}; d\theta_1^{(j)}, d\tilde{\theta}_2^{(\gamma_{j-1}+1:\gamma_j)} \right) \times \dots \\ & \dots \times \tilde{K}_n \left( \theta_1^{(n-1)}, \tilde{\theta}_2^{(1:\gamma_{n-1})}; d\theta_1^{(n)}, d\theta_2^{(\gamma_{n-1}+1:\gamma_n)} \right), \end{aligned}$$

and for  $k, j > m$

$$\begin{aligned} & \mu \tilde{K}_{1:k-1} \hat{K}_{k:n,m} \left( d\theta_1^{(n)}, d\theta_2^{(n)} \right) = \int_{\Theta_1^n \times \Theta_2^{\gamma_m-1}} \mu \left( d\theta_1^{(0)} \right) \\ & \times \tilde{K}_1 \left( \theta_1^{(0)}; d\theta_1^{(1)}, d\tilde{\theta}_2^{(1)} \right) \tilde{K}_2 \left( \theta_1^{(1)}, \tilde{\theta}_2^{(1)}; d\theta_1^{(2)}, d\tilde{\theta}_2^{(2:\gamma_2)} \right) \dots \\ & \int_{\Theta_2^{\gamma_m-\gamma_{m-1}}} \dots \int_{\Theta_2^{\gamma_j-\gamma_{m-1}}} \dots \int_{\Theta_2^{\gamma_n-\gamma_{m-1}}} \dots \\ & \dots \times \hat{K}_{j,m} \left( \theta_1^{(j-1)}, \tilde{\theta}_2^{(1:\gamma_{m-1})}; d\theta_1^{(j)}, d\theta_2^{(\gamma_{m-1}+1:\gamma_j)} \right) \\ & \dots \times \hat{K}_{n,m} \left( \theta_1^{(n-1)}, \tilde{\theta}_2^{(1:\gamma_{m-1})}; d\theta_1^{(n)}, d\theta_2^{(\gamma_{m-1}+1:\gamma_n)} \right). \end{aligned}$$

Now that notation is defined we can express the main result of this section. We want to study the asymptotic properties of the difference of the two stochastic processes, more precisely we want to prove that under certain conditions for any probabilities  $\nu$  and  $\mu$

$$\lim_{n \rightarrow +\infty} \left\| \nu K_{1:n} - \mu \tilde{K}_{1:n} \right\| = 0.$$

A trivial decomposition and the application of the triangle inequality leads to

$$\left\| \nu K_{1:n} - \mu \tilde{K}_{1:n} \right\| \leq \left\| \nu K_{1:n} - \mu K_{1:n} \right\| + \left\| \mu K_{1:n} - \mu \tilde{K}_{1:n} \right\|.$$

From the result of the previous subsection, the SAME algorithm is ergodic and thus the first term goes to zero as  $n \rightarrow +\infty$ . Consequently we focus on the second term.

Our results are based upon a decomposition into an estimation error and an approximation bias, which we now state:

**Proposition 1** For all integers  $m_n$ , and  $n$  such that  $m_n < n$ , we have the estimate

$$\begin{aligned} \left\| \mu K_{1:n} - \mu \tilde{K}_{1:n} \right\| & \leq \left\| \mu K_{1:n} - \mu \tilde{K}_{1:m_n} \hat{K}_{m_n+1:n,m_n} \right\| \\ & \quad + \sum_{k=m_n+1}^n \left\| \mu \tilde{K}_{1:k-1} \hat{K}_{k,m_n} - \mu \tilde{K}_{1:k} \right\|. \end{aligned}$$

**Proof.** For  $m_n < n$  we have the telescoping sum

$$\begin{aligned} \mu K_{1:n} - \mu \tilde{K}_{1:n} & = \mu K_{1:n} - \mu \tilde{K}_{1:m_n} \hat{K}_{m_n+1:n,m_n} \\ & \quad + \sum_{k=m_n+1}^n \mu \tilde{K}_{1:k-1} \hat{K}_{k:n,m_n} - \mu \tilde{K}_{1:k} \hat{K}_{k+1:n,m_n}, \end{aligned}$$

with the convention  $\hat{K}_{n+1:n,m_n} = Id$ . Then by first applying the triangle inequality and the fact that for any probability measures  $\mu$  and  $\nu$  the following statement holds  $\left\| \mu \hat{K}_{k,m_n} - \nu \hat{K}_{k,m_n} \right\| \leq \left\| \mu - \nu \right\|$  we obtain the result. ■

**Proposition 2** There exists a sequence  $m_n$  such that

$$\lim_{n \rightarrow +\infty} \left\| \mu K_{1:n} - \mu \tilde{K}_{1:m_n} \hat{K}_{m_n+1:n,m_n} \right\| = 0.$$

Intuitively, during the  $m_n$  first iterations  $\tilde{K}_{1:m_n}$  introduces an approximation error compared to the SAME algorithm, which is then corrected in the following  $n-m_n+1$  iterations with  $\hat{K}_{m_n+1:n,m_n}$ . Then if  $m_n$  increases significantly less fast than  $n$  such that  $\hat{K}_{m_n+1:n,m_n}$  can correct and forget in  $n-m_n+1$  iterations the error generated during the  $m_n$  first iterations, then the result should hold.

**Proposition 3** There exists a sequence  $m_n$  such that

$$\lim_{n \rightarrow +\infty} \sum_{k=m_n+1}^n \left\| \mu \tilde{K}_{1:k-1} \hat{K}_{k,m_n} - \mu \tilde{K}_{1:k} \right\| = 0.$$

This result relies on the fact that for term  $k$  in the sum, the two dynamics are the same up to time  $k-1$  and simply differ at iteration  $k$  where, on one hand, the  $\theta_2^{(m_n:k)}$  are “rejuvenated” with  $\hat{K}_{k,m_n}$  and on the other hand only  $\theta_2^{(k)}$  is sampled with  $\tilde{K}_k$ . When  $\theta_1$  and  $\theta_2$  lie in bounded spaces one can bound the error introduced, and show that there exists  $0 < \beta < 1$  such that for  $m_n = n - n^\beta$  the sum of these errors goes to zero as  $n \rightarrow +\infty$ .

By combining the three propositions and using the convergence result proved for the SAME algorithm we can deduce the following result:

**Theorem 4** There exist sequences  $m_n$  and  $\gamma_n$  such that for any  $\mu$ ,

$$\lim_{n \rightarrow +\infty} \left\| \bar{p}^{\gamma_n} - \mu \tilde{K}_{1:n} \right\| = 0,$$

which proves the validity of the SA<sup>2</sup>ME algorithm under suitable conditions. Note that these results rely on a boundedness assumption on the parameters. We are currently extending these results to more general cases for other problems.

## 6 Simulation results

We applied the two algorithms described above for the following parameters:  $T = 64$  and  $k = 2$ . We define  $E_i \triangleq a_{c_i}^2 + a_{s_i}^2$ .  $E_1 = 20$ ,  $E_2 = 6.32$ ,  $-\arctan(a_{s_1}/a_{c_1}) = 0$ ,  $-\arctan(a_{s_2}/a_{c_2}) = \pi/4$ ,  $\omega_1/2\pi = 0.2$  and  $\omega_2/2\pi = 0.3$ . The SNR is defined as  $10 \log_{10} E_1/(2\sigma^2)$  and equal to 1dB. Theoretically, the algorithms require a so-called logarithmic cooling schedule  $\gamma_i$  and an infinite number of iterations to converge. This sequence goes to  $+\infty$  too slowly to be used practically. We run here the algorithms for 500 iterations and select a linear growing cooling schedule  $\gamma_i = A + Bi$  where  $\gamma_0 = 1$  and  $\gamma_{500} = 10^2$ . We used the same series  $\gamma_i$  for the second algorithm and set  $i_0 = 20$ . Note the slower convergence of the second algorithm compared with the first one, as expected.

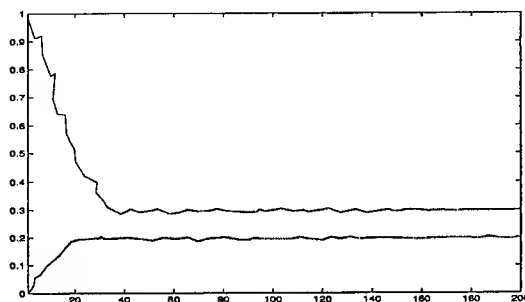


Figure 1: Convergence of the SAME towards the marginal MAP estimates of the frequencies

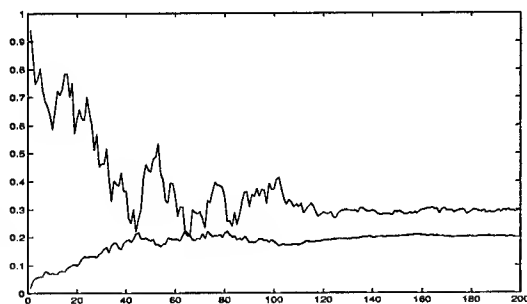


Figure 2: Convergence of the SA<sup>2</sup>ME algorithm towards the marginal MAP estimates of the frequencies

## 7 REFERENCES

- [1] C. Andrieu and A. Doucet, "Joint Bayesian Detection and Estimation of Noisy Sinusoids via Reversible Jump MCMC," *IEEE Trans. Signal Processing*, vol. 47, no. 10, pp. 2667-2676, 1999.
- [2] P. Barone, R. Ragana, "Bayesian estimation of parameters of a damped sinusoidal model by a Markov chain Monte Carlo method," *IEEE Trans. Sig. Proc.*, 45 (7) (1997) 1806-1814.
- [3] J.M. Bernardo, A.F.M. Smith, *Bayesian Theory*, Wiley series in Applied Probability and Statistics, 1994.
- [4] G.L. Bretthorst, "Bayesian Spectrum Analysis and Parameter Estimation," *Lecture Note in Statistics*, vol. 48, Springer-Verlag, New-York, 1988.
- [5] P.M. Djurić, H. Li, "Bayesian spectrum estimation of harmonic signals," *IEEE Sig. Proc. Letters*, 2 (11) (1995) 213-215.
- [6] A. Doucet and C. Andrieu, "Robust Bayesian spectral analysis using MCMC," in *Proc. EUSIPCO'98*, Island of Rhodes, Sept. 1998.
- [7] E.T. Jaynes, "Bayesian Spectrum and Chirp Analysis," in *Maximum Entropy and Bayesian Spectral Analysis and Estimation Problems*, Ed. D. Reidel, Dordrecht-Holland, 1987, 1-37.
- [8] A. Philippe, "Simulation of right and left truncated gamma distributions by mixtures," *Statistics and Computing*, 7, (1997), 173-181.
- [9] D.C. Rife, R.R. Boorstyn, "Multiple-tone parameter estimation from discrete-time observations," *Bell Syst. Tech. J.*, 55 (1976) 1389-1410.
- [10] C.P. Robert, A. Doucet and S.J. Godsill, "Marginal Maximum A Posteriori Estimation using MCMC," *Proc. IEEE ICASSP'99*.
- [11] P. Stoica, R.L. Moses, B. Friedlander, T. Soderstrom, "Maximum likelihood estimation of the parameters of multiple sinusoids from noisy measurements," *IEEE Trans. Acou. Speech Sig. Proc.*, 37 (1989) 378-392.

# HARMONIC ANALYSIS ASSOCIATED WITH SPATIO-TEMPORAL TRANSFORMATIONS.

Jean-Pierre Leduc

Washington University in Saint Louis, Department of Mathematics  
One Brookings Drive, P.O. Box 1146, Saint Louis, MO 63130  
Email: leduc@math.wustl.edu

## ABSTRACT

The paper presents new developments in harmonic analysis associated with the motion transformations embedded in digital signals. In this context, harmonic analysis provides motion analysis with a complete theoretical construction of perfectly matching concepts and a related toolbox leading to fast algorithms. This theory can be built from only two assumptions: an associative structure for the local motion transformations expressed as Lie group and a principle of optimality for the global evolution expressed as a variational extremal. Motion analysis means not only detection, estimation, interpolation, and tracking but also propagators motion-compensated filtering, signal decomposition, and selective reconstruction. The optimality principle defines the trajectory and provides the appropriate equations of motion, the selective tracking equations, the selective constants of motion to be tracked, and all the symmetries to be imposed on the system. The harmonic analysis provides new special functions, orthogonal bases, PDE's, ODE's and integral transforms. The tools to be developed rely on group representations, continuous and discrete wavelets, the estimation theory (prediction, smoothing and interpolation) and filtering theory (Kalman filters, motion-based convolutions, integral transforms). All the algorithms are supported by fast and parallelizable implementations based on the FFT and dynamic programming.

## 1. INTRODUCTION

In this paper, the harmonic analysis on motion transformations is built on the actual kinematics as they take place in the external space and in the projections on sensor arrays (Figure 1). Eventually, they are embedded in the signals to analyze. From that point of view, this approach fundamentally differs from the motion models currently presented in the Literature (see in [1] and all the references) which rely on techniques based on stochastic processes, statistics and operations research. These are namely block-matching, pel-recursive and Bayesian techniques. As a major drawback, these techniques are totally blind to the underlying mathematical structures of the spatio-temporal transformations.

The author wants to thank Prof. B. Blank in the Math. Dept. for helpful discussions and Prof. B. K. Ghosh in the SSM Dept. for his support on numerical computations. This research work was supported by the AFOSR grant No. F49620-99-1-0068.

The main point of the approach proposed in this paper is to bring differential geometry, mechanics on manifold and harmonic analysis into signal analysis. This theory provides the actual kinematics and relies only on two key assumptions that can be summarized as follows: a Lie group structure (i.e. an associative law of composition, and an identity element for the local transformations) and a principle of optimality (for the global evolution). From those two key points, a complete machinery of theory, analysis tools and fast algorithms can be constructed in such a nice way that all the concepts perfectly match to each other. This paper presents new developments on this important topic that cover all the kinematics embedded in any spatio-temporal real and complex signals and apply to video, radar and sonar.

The construction of Lie group representations (i.e. the analyzing functions in the signal space) leads naturally to several important topics. First, this leads to the existence of *continuous wavelet transforms with frames, tight frames and new discrete wavelets* placed along the trajectories which perform spatio-temporal and motion-based atomic decompositions, expansions, filtering (prediction, smoothing and interpolation), estimation and motion-selective reconstructions. The second topic deals with *the characters* of the group representations to define new special functions and integral transforms (IT) which generalize the Fourier kernel for the new kinematics of interest. The third topic proceeds with the *adjunction of a principle of optimality* based on Euler-Lagrange equations and define the existence of a trajectory and a tracking. This gives rise to the Partial Differential Equations (PDE) as equations of wavelet and signal motion and to Ordinary Differential Equations (ODE) for tracking. Fourth, the Green functions associated with these PDE's turn out to be the previous special functions related to the kinematics. At this stage, we yield a global analysis structure with the construction of signal propagators, and motion-compensated filters.

## 2. GROUP REPRESENTATIONS, WAVELETS AND CONVOLUTIONS

In their general form, the Lie group representations  $\hat{T}_g$  acting upon functions  $\hat{\Psi} \in L^2(\mathbb{R}^n \times \mathbb{R}, d\vec{k}d\omega)$  read

$$[\hat{T}_g \hat{\Psi}](\vec{k}, \omega) = a^{n/2} e^{i(\omega\tau + \vec{k} \cdot \vec{b})} \hat{\Psi}[g^{-1}(\vec{k}, \omega)] \quad (1)$$

where  $g$  is an element of the group  $G$ , the  $L^2$  normalizing factor  $a^{n/2}$  originates from a Radon-Nikodym derivative and provides unitary representations,  $e^{i(\omega\tau + \vec{k}\cdot\vec{b})}$  stands for the character of the subgroup of spatio-temporal translations, and  $g^{-1}(\vec{k}, \omega)$  is the left-group action of  $g \in G$  in the dual space. The dual space, also called the phase space, is the Fourier domain denoted  $\hat{\cdot}$  with spatial frequencies  $\vec{k} \in \mathbb{R}^n$  and temporal frequency  $\omega \in \mathbb{R}$ . The parameters  $a \in \mathbb{R}^+ \setminus \{0\}$ ,  $\vec{b} \in \mathbb{R}^n$ , and  $\tau \in \mathbb{R}$  are respectively the scale, the spatial and temporal translations.

From the group representations, we define the *continuous wavelet transform* as the operator  $W_\psi$  mapping the function  $S \in H = L^2(\mathbb{R}^n \times \mathbb{R})$  into functions of  $g$  defined as

$$[W_\psi S](g) = \int_{\mathbb{R}^n \times \mathbb{R}} S(\vec{x}, t) \overline{[T_g \Psi](\vec{x}, t)} d^n \vec{x} dt = \langle S | T_g \Psi \rangle \quad (2)$$

This inner product  $\langle \cdot, \cdot \rangle$  would remain a simple correlation between  $[T_g \Psi](\vec{x}, t)$  and  $S(\vec{x}, t)$  if no further conditions were imposed on the unitary and irreducible group representations. In fact, to be a continuous wavelet transform, the mapping must be invertible i.e. that there exists an operator  $W_\psi^{-1}$  such that  $W_\psi^{-1} W_\psi = I_H$ .  $I_H$  is the identity operator in the Hilbert space of observation  $H$ . This means that we want to perfectly reconstruct the signal

$$S(\vec{x}, t) = \int_G [W_\psi S](g) [T_g \Psi](\vec{x}, t) d\lambda_l(g) \quad (3)$$

$d\lambda_l$  is the left-invariant Haar measure calculated on the group  $G$ . The condition to be fulfilled in order to derive the inverse transform is known since 1964 in the work of Calderón. Several examples considered in this paper are defined in [4, 5, 6, 7]. The simplest case is the affine-Galilean group where the group element is  $g = \{\vec{b}, \tau, \vec{v}, a\}$  where  $v \in \mathbb{R}^n$  is the velocity vector [6]. The left-group action is given by  $(\frac{1}{a}[\vec{x} - \vec{b} - \vec{v}(t - \tau)], t - \tau)$  and the representation in Equation (1) reads  $[\hat{T}_g \hat{\Psi}](\vec{k}, \omega) = a^{n/2} e^{i(\omega\tau + \vec{k}\cdot\vec{b})} \Psi[\vec{k}', \omega']$  with  $\vec{k}' = a\vec{k}$ ,  $\omega' = \omega + \vec{k} \cdot \vec{v}$ . Let us examine the condition for an invertible transform in the affine-Galilean case with  $n = 1$  i.e.  $b, \tau, v \in \mathbb{R}$  and  $a \in \mathbb{R}^+ \setminus \{0\}$  as follows

$$F(\vec{x}, t) = \int_G \langle F | T_g \Psi \rangle (T_g \Psi)(X) d\lambda_l(g) \quad (4)$$

which becomes after some easy computations

$$= \int_{\mathbb{R}} \int_{\mathbb{R}^+} \left\{ \int_{\mathbb{R} \times \mathbb{R}} F \left( \begin{array}{c} y \\ \rho \end{array} \right) (\Psi_a *_{\vec{v}} \tilde{\Psi}_a) \left[ \begin{array}{c} (x-y) - v(t-\rho) \\ t-\rho \end{array} \right] dy d\rho \right\} \frac{dv da}{a^2} \quad (5)$$

$$= \int_{\mathbb{R}} \int_{\mathbb{R}^+} \left\{ F *_{\vec{v}} (\Psi_a *_{\vec{v}} \tilde{\Psi}_a) \right\} \left( \begin{array}{c} x \\ t \end{array} \right) \frac{dv da}{a^2} \quad (6)$$

where we have let  $\tilde{\Psi}(x, t) = \overline{\Psi(-x, -t)}$  and  $\Psi_a(x, t) = a^{-1} \Psi(\frac{x}{a}, t)$ . Let us make an important remark about Equations (5) and (6): the introduction of a non-conventional spatio-temporal convolution denoted  $*_{\vec{v}}$  is in fact a convolution twisted along the Galilean transformation i.e. the translation in space has a component depending on time

and moving at the constant velocity  $v$ . The convolution performed along this displacement (i.e. along the trajectory) allows the reconstruction of the still signal  $F(x, t)$ . This property is in fact a reminiscence of the motion-compensated filtering developed by the author in [3] which is going to be generalized in this work by the introduction of IT's. Eventually, let us move to the Fourier domain and retrieve the usual condition of admissibility for the Galilean wavelet as described in [9, 12, 13]. Proceeding with Equation (5) in the Fourier domain, we obtain

$$\hat{F}(\vec{k}, \omega) = \hat{F}(\vec{k}, \omega) \int_{\mathbb{R}} \int_{\mathbb{R}^+ \setminus \{0\}} |\hat{\Psi}(a\vec{k}, \omega - \vec{k} \cdot \vec{v})|^2 \frac{dv da}{a^2} \quad (7)$$

which leads to the usual condition of square-integrability of the Galilean wavelet in one-dimensional space and time

$$\int_{\mathbb{R}} \int_{\mathbb{R}} \frac{|\hat{\Psi}(k, \omega)|^2}{|k|^2} dk d\omega = 1 \quad (8)$$

See references [4, 5, 6, 7] for the properties and applications of the Galilean wavelets.

The construction of orthonormal bases proceeds by discretizing the group parameters into a lattice. The spatio-temporal lattice  $\mathfrak{z}$  is easily defined as a generalization of the discretization affine group  $a = a_*^m$ ,  $b = n_b b_* a_*^m$ ,  $v = n_v v_* a_*^m$ ,  $\tau = n_\tau \tau_*$  with  $a_* > 1$ , and  $b_*, v_*, \tau_* > 0$  for convenience. If we now consider the regular left-composition  $g^{-1}(x, t) = (\frac{x - b - v(t - \tau)}{a}, t - \tau)$  in the Galilean case, we can mimic the case of the affine group [6] as follows. Let  $a_* = 2$  and  $T_{g_*} \Psi(x, t) = a_*^{-m/2} \Psi(a_*^{-m} x - n_b b_* - n_v v_*(t - n_\tau \tau_*), t - n_\tau \tau_*)$  where we retrieve the well-known orthonormal bases  $\Psi_{m,p,q}(x, t) = 2^{-m/2} \Psi(2^{-m} x - p, t - q)$  in  $L^2(\mathbb{R} \times \mathbb{R})$  at  $p = n_b b_* + n_v v_* n_\tau \tau_*$ ,  $q = n_\tau \tau_*$  with  $p, q \in \mathbb{Z}$ . Technically, we have deployed the usual discrete wavelets defined from the affine group along spatio-temporal translations that correspond to the motion trajectories at constant velocity [3].

### 3. SPECIAL FUNCTIONS AND INTEGRAL TRANSFORMS

In this section, we proceed one step further on the representations and focus on the characters. The integration of the characters leads to special functions. These special functions naturally define the kernel of an integral transform. This procedure can be performed for each group of spatio-temporal transformation. Let us consider an important example known as rotational motion (described in [5]). The set of parameters is given as  $G = \{g | g = (\vec{b}, \tau, \theta_1, a)\}$  where  $\theta_1$  is the angular velocity  $\theta_1 \in \mathbb{R}$ . The composition law is given as  $g \circ g' = \{\vec{b} + aR(\theta_1\tau)\vec{b}', \tau + \tau', \theta_1 + \theta_1', aa'\}$ ; the inverse element reads as  $g^{-1} = \{-a^{-1}R(\theta_1\tau)^{-1}\vec{b}, -\tau, -\theta_1, a^{-1}\}$ . The group representations  $[\hat{T}(g)\hat{\Psi}](\vec{k}, \omega)$  in polar coordinates  $\vec{b} \rightarrow (r, \theta_b)$  and  $\vec{k} \rightarrow (k, \theta_k)$  with  $n = 2$  read

$$\theta_1^{\frac{1}{2}} a^{\frac{1}{2}} i^{\Omega} e^{i\Omega(\theta_k + \theta_b)} e^{i(\Omega\chi + k\tau \sin[\chi])} \hat{\Psi}(ak, \theta_k, \theta_1, \Omega) \quad (9)$$

with  $\chi = \theta_b - \theta_k + \theta_1 \tau$  and  $\Omega = \frac{\omega}{\theta_1}$ . The characters of this representation lead to the special functions (Figure 4)

$$J_\Omega(k) = \frac{1}{2\pi} \int_0^{2\pi} e^{i[\Omega u + k \sin u]} du \quad (10)$$

which are usually NOT Bessel functions except when  $\theta_1$  takes an integer values. The complexification of  $u \rightarrow i y$  gives rise to hyperbolic motion instead of circular rotations along with new special functions as in (10) with instead real exponential and *sinh* functions. These special functions can be also easily obtained by considering  $\Psi$  as a Dirac measure and integrating this measure along the trajectory. This process corresponds to "mechanics of moving points" and defines the spectral signatures of objects moving according to such transformation. The usual way to deduce the ODE which admits this special function as solution is to calculate the Laplace-Baltrami differential operator on this group. Theorems of additivity for these special functions can be deduced from the composition of the translations. In this case, it reads

$$\int_{-\infty}^{+\infty} J_{[\Omega_1 - t]}(k r_1) J_{[t - \Omega_2]}(k r_2) dt = J_{[\Omega_1 - \Omega_2]}(k(r_1 + r_2)) \quad (11)$$

Equation 10 leads to "Hankel-like" integral transforms

$$[H_\Omega f](k) = \int_0^\infty f_r(r) J_\Omega(k r) r^{n-1} dr \quad (12)$$

The same procedure and computations can be done on all the groups dealing with spatio-temporal transformations defined in [4, 5, 6, 7]. Examples on the Galilean group [6, 7] proceed with

$$\int_{\mathbb{R}} e^{-i\omega\tau} e^{-ik[x-v(t-\tau)]} d\tau = \delta(\omega + kv) e^{ik(x-vt)},$$

on the acceleration group [4] with

$$\int_{\mathbb{R}} e^{-i\omega\tau} e^{-ik[x+\frac{\gamma_2}{2}(t-\tau)^2]} d\tau = e^{i\frac{\pi}{4}} \sqrt{\frac{2\pi}{k\gamma_2}} e^{ikb} e^{ik\frac{\gamma_2}{2}\tau^2} e^{-i\frac{(\omega-\tau k\gamma_2)^2}{2k\gamma_2}}$$

where  $\gamma_2 \in \mathbb{R}$ , on the deformations [8] with

$$\int_{\mathbb{R}} e^{-i\omega\tau} e^{-ike^{s_1}t} d\tau = \frac{1}{s_1} \Gamma(-i\frac{\omega}{s_1}) e^{-i\omega\frac{\log(i\pi k)}{s_1}}$$

where  $s_1 \in \mathbb{R}$  and  $\Gamma()$  is the usual Gamma function.

#### 4. PRINCIPLE OF OPTIMALITY AND TRACKING

According to calculus of variations, the motion between times  $t_1$  and  $t_2$  coincides with the extremal of the functional  $J$

$$\delta J = 0 \quad \text{with } J = \int_{t_1}^{t_2} L[\vec{q}(t), \dot{\vec{q}}(t), \ddot{\vec{q}}, \dots, \vec{q}^{(k)}; t] dt, \quad (13)$$

where  $\delta$  stands for the variation. The application of the optimal variational principle in Equation (13) is equivalent to writing the so-called Euler-Lagrange equation [7]. The trajectory is then uniquely defined if the initial state  $\vec{q}(0) = \vec{q}_0$  of the object is known. At the extremum, denoted by the subscript  $*$ , the Euler-Lagrange equation

$$\frac{d}{dt} \frac{\partial L}{\partial \dot{\vec{q}}_*} - \frac{\partial L}{\partial \vec{q}} = 0. \quad (14)$$

This Euler-Lagrange equation generalizes quite easily and, moreover, allow us to derive the *equation of wavelet motion* that optimizes the action  $J$ . If we consider the Galilean case with one-dimensional space with  $q(\tau) = b(\tau)$  and  $\dot{q}(\tau) = \dot{b}(\tau)$  and the inner product 2 as Lagrangian, then (14) becomes

$$\frac{d}{d\tau} \frac{\partial <\Psi_{[b, \dot{b}, \tau]} | S >}{\partial \dot{b}} - \frac{\partial <\Psi_{[b, \dot{b}, \tau]} | S >}{\partial b} = 0. \quad (15)$$

It is convenient to expand the total differential. The conditions to introduce the operator in the integral are fulfilled. One solution of this IT is that the kernel be equal to 0. This gives a PDE on  $\hat{\Psi}(ak, \omega - k\dot{b})$  i.e. the motion equation for the wavelet. In the Fourier domain the PDE operator  $\hat{\Lambda}_{(\dot{b}, \ddot{b}, k, \omega)}$  is given by

$$(\dot{b}k + \omega) \left( \frac{\partial}{\partial \omega} - \frac{1}{\dot{b}} \frac{\partial}{\partial k} \right) - i \ddot{b} \left[ \frac{1}{\dot{b}^2} \frac{\partial}{\partial k} - k \left( \frac{\partial}{\partial \omega} - \frac{1}{\dot{b}} \frac{\partial}{\partial k} \right)^2 \right] \quad (16)$$

and the PDE by  $\hat{\Lambda}_{(\dot{b}, \ddot{b}, k, \omega)} \hat{\Psi}(ak, \omega - k\dot{b}) = \hat{\Psi}(ak, \omega - k\dot{b})$ . There are many applications out of this procedure which can be similarly drawn for each spatio-temporal group. Two are examined below and a third in Section 5.

If we consider a wavelet tuned on parameter  $g_1$  and the Dirac measure on parameter  $g_2$ , the partial differential operator  $\hat{\Lambda}_{(\dot{b}, \ddot{b}, k, \omega)}$  becomes  $\hat{\Pi}_{(v_1, v_2, \dot{v}_1; k, \omega)}$

$$(v_1 k + \omega) \left( \frac{1}{v_2 - v_1} \frac{d}{dk} \right) + i \dot{v}_1 \left[ \frac{1}{(v_2 - v_1)^2} \frac{d}{dk} - k \frac{1}{v_2 - v_1} \frac{d^2}{dk^2} \right] \quad (17)$$

and the PDE becomes an ODE i.e. the tracking equation  $\hat{\Pi}_{(v_1, v_2, \dot{v}_1; k, \omega)} \hat{\Psi}((ak, -k(v_2 - v_1))) = \hat{\Psi}(ak, -k(v_2 - v_1))$ .

Let us consider a Galilean Morlet wavelet [6, 7] applied to a Dirac measure in pure translation motion at constant velocity. The signal taken as a Dirac measure on a translational trajectory is given by  $S(x, t) = \delta[x - vt]$ . The Lagrangian  $\Phi[b, \tau, v; k_0, \omega_0] = <\Psi_g | S >$  reads after integrating the inner product, we get  $\Phi[b, \tau, v; k_0, \omega_0] =$

$$\sqrt{2\pi} e^{ik_0(b - \dot{b}\tau)} e^{-\frac{1}{2}\{(b\tau - b)^2 + \tau^2\}} e^{-i\omega_0\tau} e^{\frac{1}{2}\{(k_0 v - k_0 \dot{b} + \omega_0)^2 - ((v - \dot{b})(b\tau - b) - \tau)^2\}} e^{i\{(k_0 v - k_0 \dot{b} + \omega_0)((v - \dot{b})(b\tau - b) - \tau)\}} \quad (18)$$

$k_0$  and  $\omega_0$  are the coordinates of the wavelet shift in the Fourier domain. The contribution of all the partial derivatives involved in the Euler-Lagrange equation namely leads to an ODE in form of a product of  $F$ , which is a complex function of the constant of motion  $\dot{b}\tau - b$  and  $2\dot{b} = -\dot{b}\tau$ , with the Lagrangian  $\Phi[b, \tau, v; k_0, \omega_0]$

$$F[\dot{b}\tau - b, \ddot{b}\tau - 2\dot{b}] \Phi[b, \tau, v, k_0, \omega_0] = 0 \quad (19)$$

such that  $F(0, 0) = 0$ . The ODE vanishes when  $v = \dot{b}$ ,  $\dot{b}\tau - b = 0$ ,  $\ddot{b}\tau - 2\dot{b} = 0$  and  $\omega_0 = 0$ ,  $k_0 \neq 0$ . Therefore, we have verified that the tracking addresses the correct constant of motion  $b = \dot{b}\tau$  and  $b = \dot{b}\tau + \frac{1}{2}\dot{b}\tau^2$  meaning that the system can track objects at constant velocity and constant second-order acceleration. The tracking requires some symmetry in the wavelet i.e. that the still wavelet must be located in the plane  $\omega = 0$  with  $k_0 \neq 0$ . These practical results have algorithmic importance as pictured in [7].

## 5. MOTION-BASED FILTERING

This section extends the concept of velocity filtering originally defined by Fleet and Jepson in [1], studied by Dubois in [2] for all the categories of motion within the approach pursued in the previous section. To reach that goal, we introduce integral transforms whose kernels are motion-specific Green functions. In the following, it is demonstrated that the motion-specific Green functions can be equivalently derived from the characters of the group representations in Section (3) or from the fundamental solution of the PDE of the wave equation of Section (4). This leads to convolutional integral transforms twisted along the motion transformations as presented in Section (2). The interesting point of this approach comes from the Equations of the wavelet motion themselves (16) expressed in the Fourier domain. As a result of the existence of the term  $\frac{\partial L}{\partial b}$  in  $\hat{\Lambda}$ , the PDE can be re-written in the Fourier domain in the form of

$$\hat{\Lambda} \hat{\Psi}(g^{-1}X) = \hat{\Psi}(g^{-1}X) \quad \text{where } X = (\vec{k}, \omega) \quad (20)$$

with an eigen value at 1. The Green function  $G$  for operator  $\hat{\Lambda}$  is the distribution which satisfies  $\hat{\Lambda} \hat{G}(g^{-1}X) = \hat{\delta}(g^{-1}X)$ . The Green function is the Dirac  $\delta(g^{-1}X)$  itself. The Green function known as the fundamental solution of the PDE as in Equation (20). If the operator  $\hat{\Lambda}$  is injective then, the inverse  $\hat{\Lambda}^{-1}$  exists and provides a convolutional-type integral transforms whose kernel is the Green function i.e.

$$[G(f)](\xi) = \int_a^b G(x, \xi) f(x) dx \quad (21)$$

These kernels are meaningful and remind the propagators associated with Green functions of the Schrödinger equations. The meaning of Equation (21) and of the wavelet-based reproducing kernels [7] leads to the following duality of the motion analysis.

(1.) If the still version of a signal (wavelet, filter or stochastic process)  $f(x)$  is known, then reproducing kernel integral transform provides all the moving version in  $(x, t)$  or in  $(k, \omega)$ . These integral transforms generate the whole family of analyzing signals, wavelets or processes in the observing space  $L^2(\mathbb{R}^n \times \mathbb{R}, d^n \vec{x} dt)$ . This allows spatio-temporal filtering, interpolating, and predicting along a trajectory.

(2.) If the animated version of a signal is known, then Equation (21) is a filter that compensates the signal from a given motion and gives rise to the still signal. This is motion-compensation filtering. The advantage of such approach is that the classical affine wavelet analysis and processing may then be applied on the compensated signal (for coding purpose as in [3]). This section brings a more general point of view on the motion analysis presented in [3] where motion compensated filtering was performed by building the trajectories within the signal and applying discrete wavelets along the assumed trajectories.

Let us then revisit Section (3) and compute the Fourier transform of a Dirac measure on a trajectory

$$K(\vec{k}, \omega; m; \vec{x}, t) = \int_{\mathbb{R}} \exp[-i(\omega\tau + \vec{k} \cdot (g^{-1}\vec{x})|_{x=0})] d\tau \quad (22)$$

If  $g = e$  the identity element, we retrieve the usual Fourier transform with kernel  $K(\vec{k}, \omega) = \delta(\omega) e^{i\vec{k} \cdot \vec{x}}$ . This procedure defines for each kind of motion the kernel  $K(\vec{k}, \omega; m; \vec{x}, t)$  that particularizes the usual Fourier transform for the motion group of interest.  $m$  denotes the current motion parameter. If the Dirac measure is transformed into a continuous wavelet with compact support, then the calculation of  $\hat{\Psi}(\vec{k}, \omega; m)$  animated of motion  $m$  from its still cognate  $\Psi(\vec{x}, t)$  becomes an integral transform with kernel  $K$ . Let us, for example, consider the kernel of accelerated wavelets as propagator presented in section (3) and integrate with a still Morlet wavelet [6, 7], this yields the propagated wavelets for second-order accelerations

$$\hat{\Psi}_{\vec{\gamma}_2}(\vec{k}, \omega) = (2\pi) e^{i\frac{\pi}{4}} \sqrt{\frac{2\pi}{\vec{k} \cdot \vec{\gamma}_2}} e^{-i\frac{1}{2} \frac{\omega^2}{\vec{k} \cdot \vec{\gamma}_2}} e^{-\frac{(k-k_0)^2}{2}} e^{-\frac{\omega^2}{2}} \quad (23)$$

Moreover, as the function  $\Psi$  can now be scaled to extend the results from the "point mechanics" towards the "object-based mechanics" as follows

$$\hat{\Psi}(\vec{k}, \omega; m, a, a_0) = \int_{\mathbb{R} \times \mathbb{R}^n} K(\vec{k}, \omega; m; \vec{x}, t) \Psi(\vec{x}, t; a, a_0) d^n \vec{x} dt \quad (24)$$

We have reach so far the ability to generate, cancel or modify analyzing wavelets as well as moving patterns.

## 6. CONCLUSIONS AND APPLICATIONS

This paper has shed light on a novel motion analysis based on a group-theoretic approach. Let us consider the projection of moving patterns on sensor arrays which creates the most important part of all the acceleration components embedded in signals. The traffic sequence (Figure 2) is an example. The projection takes place within the cone of sensor visibility (Figure 1) is a homothety (i.e. a re-scaling). The projection may be modelled as an orthogonal projection composed with a scaling. Let us define the  $z$ -axis orthogonal to the sensor plane and the  $x-y$  axes in the sensor plane. The motion captured in the sensor plane is obtained after a projection on planes  $\Pi_0, \Pi_1, \Pi_2$  parallel to sensor at time  $\tau = 0, 1, 2$  and a homothety that rescales the projection down to the plane of the sensor (Figure 1). Let us denote  $W$  the width of the rigid object and  $S_0$  the size of the object captured by the camera. The scale  $a_0 = \frac{W}{S_0}$  is observed from plane  $\Pi_0$  at time  $\tau = 0$ . At time  $\tau = n$ , the size perceived from plane  $\Pi_n$  by the camera is given by  $a_n = \frac{W}{S_n} = \frac{W}{S_0(1 - \frac{v_z}{d}\tau)} = \frac{a_0}{1 - \frac{v_z}{d}\tau} = a_0 [1 + \frac{v_z}{d}\tau + (\frac{v_z}{d})^2 \tau^2 + \dots] = a_0 [1 + a_1 \tau + a_2 \tau^2 + \dots + a_n \tau^n + \dots]$ . The series is convergent if  $|\frac{v_z}{d}\tau| < 1$  i.e. with the physical observation. The components of translation, velocity and accelerations along  $x$  and  $y$  axis are rescaled with the ratio  $\frac{b}{b_0} = \frac{v}{v_0} = \frac{d - v_z \tau}{d}$ .

## References

- [1.] A. Tekalp. "Digital Video Processing", Prentice-Hall, 1995.
- [2.] E. Dubois. "Motion-Compensated Filtering of Time-Varying Image", *Multidim. Syst. Sig. Proc.*, Vol. 3, pp. 211-239, 1992.

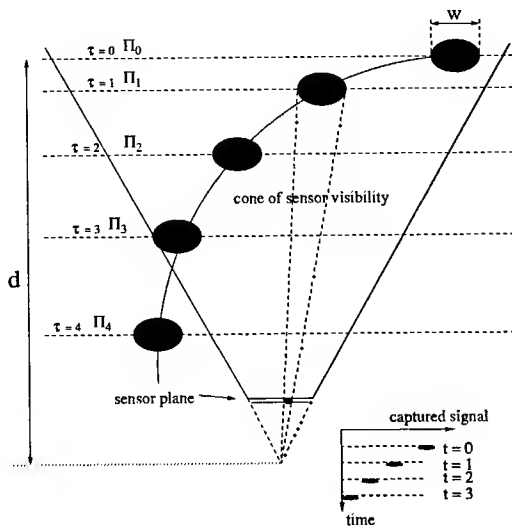


Figure 1: Tracking in a sensor cone.

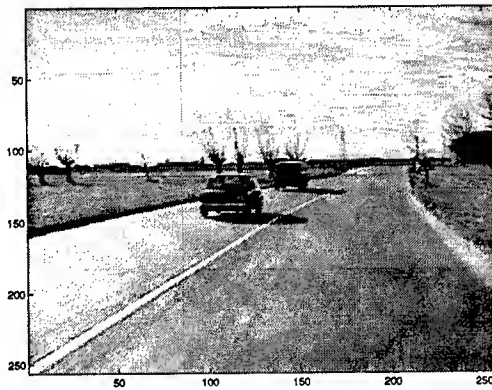


Figure 2: The 20th image of the car digital image sequences.

- [3.] J.-P. Leduc, J.-M. Odobez and C. Labit. "Adaptive Motion-Compensated Wavelet Filtering for Image Sequence Coding", *IEEE Transactions on Image processing*, Vol. 6, No. 6, pp. 862-878, June 1997.
- [4.] J.-P. Leduc, J. Corbett, M. Kong, V. M. Wickerhauser, B. K. Ghosh. "Accelerated Spatio-temporal Wavelet Transforms: an Iterative Trajectory Estimation", *IEEE ICASSP*, Vol. 5, 1998, pp. 2777-2780.
- [5.] M. Kong, J.-P. Leduc, B. Ghosh, J. Corbett, V. Wickerhauser. "Wavelet based Analysis of Rotational Motion in Digital Image Sequences", *ICASSP-98*, Seattle, May 12-15, 1998, pp. 2781-2784.
- [6.] J.-P. Leduc, F. Mujica, R. Murenzi, M. J. S. Smith. "Spatio-Temporal Wavelet Transforms for motion tracking", *ICASSP-97*, Munich, Vol. 4, pp. 3013-3017, 1997.
- [7.] J.-P. Leduc, F. Mujica, R. Murenzi, and M. Smith. "Spatio-Temporal Wavelets: a Group-Theoretic Construction for Motion Estimation and Tracking", to appear in *SIAM Journal of Applied Mathematics*.
- [8.] J. Corbett, J.-P. Leduc, M. Kong. "Analysis of Deformational Transformations with Spatio-Temporal Continuous Wavelet Transforms", *ICASSP-99*, March 15-19, 1999.

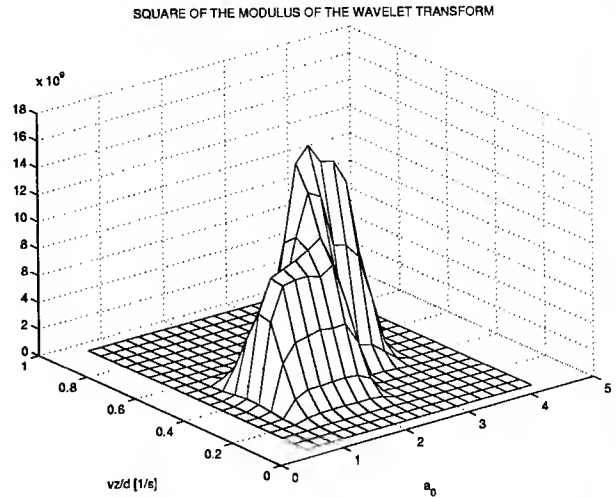


Figure 3: Estimation of the parameters  $\frac{v_z}{d}$  and  $a_0$  by computing the square modulus (energy) of the wavelet transform as in [8]:  $|\langle T(g)\Psi|s \rangle|^2 = F(a_0, \frac{v_z}{d})$  is estimated in the scene displayed in Figure 2. Two local maxima are detected and displayed at  $(\frac{v_{z1}}{d} = 0.5s^{-1}, a_{01} = 2.6)$  and  $(\frac{v_{z2}}{d} = 0.38s^{-1}, a_{02} = 1.8)$  standing for the fore and background car respectively. If we assume  $d_1 = 40$  m for the foreground car and a rate of 25 images per second, then we can estimate the relative approaching velocity component at  $v_{z1} = 72$  km/h (45 miles/h). For the background car, if we assume  $d_2 = 50$  m, then  $v_{z2} = 68.4$  km/h (42.7 miles/h). Let us remark that the camera is traveling towards the cars; therefore, both velocities correspond to relative values.

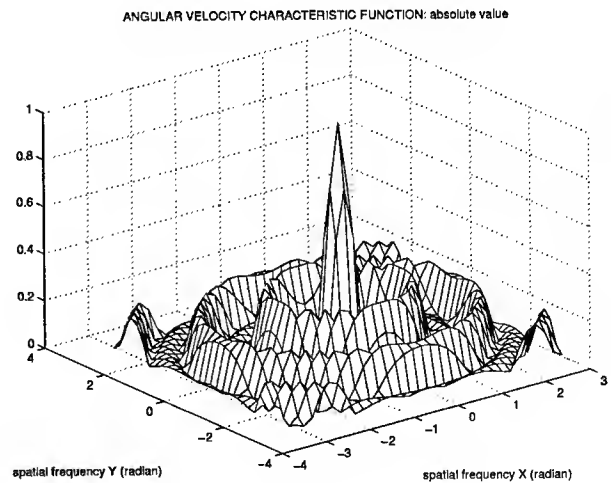


Figure 4: Spatio-temporal special function associated with the rotational motion. The sketch is performed on sections at constant  $\omega$ , the angular velocity = 1.5 radian/image.



# BLIND NOISE AND CHANNEL ESTIMATION

*M. Frikel, W. Utschick, and J. Nossek*

Technical University of Munich  
Institute for Network Theory and Signal Processing  
Arcisstr. 21, D-80290 Munich, Germany  
mifr@nws.e-technik.tu-muenchen.de

## ABSTRACT

In the classical methods for blind channel identification (Subspace method, TXK, XBM) [1, 2, 3], the additive noise is assumed to be spatially white or known to within a multiplicative scalar. When the noise is non-white (colored or correlated) but has a known covariance matrix, we can still handle the problem through prewhitening. However, there are no techniques presently available to deal with completely unknown noise fields. It is well known that when the noise covariance matrix is unknown, the channel parameters may be grossly inaccurate. In this paper, we assume the noise spatially correlated, and we apply this assumption for blind channel identification. We estimate the noise covariance matrix without any assumption except its structure which is assumed to be a band-Toeplitz matrix. The performance evaluation of the developed method and its comparison to the modified subspace approach (MSS) [4] are presented.

## 1. INTRODUCTION

One common problem in signal transmission through any channel is the additive noise. In general, additive noise is generated internally by components such as resistors, and solid-state devices used to implement the communication system. This is sometimes called thermal noise or Johnson noise. Other sources of noise and interference may arise externally to the system, such as interference from the other users. When such noise and interference occupy the same frequency band at the desired signal, its effect can be minimized by proper design of the transmitted signal and its demodulator at the receiver. The effects of noise may be minimized by increasing the power in the transmitted signal. However, equipment and other practical constraints limit the power level in the transmitted signal [5].

This work is supported by Alexander von Humboldt-Stiftung, Bundesrepublik Deutschland.

The classical model used in communication systems supposes on the one hand that the power of the noise is identical on each sensor, and on the other hand that there is no noise space/time correlation. However, this situation is seldom met, which involve a clear degradation of the performances of the subspace methods. Here, we recall some well-known methods which treat the noise problem in array processing for direction-of-arrival estimation. In fact, in recent years, there has been a growing interest in the problem of techniques with the objective of decreasing the signal to noise ratio resolution threshold or the spatially colored noise [6, 7, 8, 9, 10]. The ambient noise is unknown in practice, therefore modeling or its estimation are necessary. The methods developed for this problem are very few and there are no definitive solution. There are some practical methods; in [11] two methods are obtained by optimization of criterion and by using AR or ARMA modeling of noise. In [7] the spatial correlation matrix of noise is modeled by the known Bessel functions. As in [6] the ambient noise covariance matrix is modeled by a sum of hermitian matrices known up to multiplicative scalar. In [8] this estimate is obtained by measuring the array covariance matrix when no signals are present. This procedure assumes that the noise is not changing in function of time, which is not fulfilled in several domain applications. Another possibility [8] arises when the correlation structure is known to be invariant under a translation or rotation. The so-called differencing covariance technique can be then applied to reduce the noise influence. In this method, two identical translated and/or rotated measurements of the array covariance matrix are required and assumes the invariance of the noise covariance matrix, while the source signals change between the two measurements. The estimate noise covariance matrix is eliminated by a simple subtraction. Furthermore, this method cannot be applied when the source covariance matrix satisfies the same invariance property or when only one measurement is

available. In [7] a particular modeling structure noise covariance matrix, which takes into account the characteristic noise relative to its origins, is given. Recently, a maximum posteriori approach (MAP) has been developed in [10]; this method can only be applied in the case of a linear array. In [9], the method called "Instrumental Variable" (IV) is used to reduce the noise without estimated it; this estimator considers that the noise is temporally independent. One technique based to the *MDL* criterion has been developed in [12] for detection and localization of the signals in the presence of unknown noise; this estimator is asymptotically biased [12]. However, the study of the noise for blind channel identification is very limited. In [4], a modified subspace method (MSS) for blind identification in the presence of unknown correlated noise has been presented, indeed one use some matrices, for a time lag when the noise is absent. The object of this correspondence is to improve the blind channel identification in the presence of a correlated noise by whitening the received data. The noise is assumed spatially correlated. The structure of the paper is as follows. In the section II, we present the studied problem and in section III, we describe the noise covariance matrix model used in this study and its estimation by the proposed algorithm, we apply the noise estimation for blind channel identification using the subspace method. We present, in the section IV, some simulation results and performance comparisons.

## 2. PROBLEM FORMULATION

Consider  $L$  FIR channels driven by a common source. The output vector of the  $i$ th channel can be written as:

$$\mathbf{r}_i(k) = \mathcal{H}^{(i)} \mathbf{s}(k) + \mathbf{n}_i(k), \quad (1)$$

where,  $\mathbf{r}_i(k)$  is the output sequence of the  $i$ th channel,  $\mathbf{s}_i(k)$  is the input sequence and  $\mathbf{n}_i(k)$  is the noise sequence on the  $i$ th channel.

$$\begin{aligned} \mathbf{r}_i(k) &= [r_i(k) \quad r_i(k+1) \quad \dots \quad r_i(k+N-1)], \\ \mathbf{s}(k) &= [s(k-M) \quad s(k-M+1) \quad \dots \quad s(k+M-1)], \\ \mathbf{n}_i(k) &= [n_i(k) \quad n_i(k+1) \quad \dots \quad n_i(k+N-1)]. \end{aligned}$$

$$\mathcal{H}^{(i)} = \begin{pmatrix} h_0^{(i)} & h_1^{(i)} & \dots & h_M^{(i)} & \dots & \dots & 0 \\ 0 & h_0^{(i)} & h_1^{(i)} & \dots & h_M^{(i)} & \dots & 0 \\ \vdots & \dots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & 0 & h_0^{(i)} & h_1^{(i)} & \dots & h_M^{(i)} \end{pmatrix},$$

where,  $h_k^{(i)}$  is the impulse response of the  $i$ th channel,  $M$  is the maximum order of the  $L$  channels and  $N$  is the width of the temporal window.  $\mathcal{H}^{(i)}$  is of dimension  $(N \times (N+M))$ .

Then we have,

$$\mathbf{r}(k) = \mathcal{H} \mathbf{s}(k) + \mathbf{n}(k), \quad (2)$$

$$\begin{pmatrix} \mathbf{r}_1(k) \\ \vdots \\ \mathbf{r}_L(k) \end{pmatrix} = \begin{pmatrix} \mathcal{H}_1 \\ \vdots \\ \mathcal{H}_L \end{pmatrix} \mathbf{s}(k) + \begin{pmatrix} \mathbf{n}_1(k) \\ \vdots \\ \mathbf{n}_L(k) \end{pmatrix}.$$

The matrix  $\mathcal{H}$  is known as the  $(LN \times (N+M))$  filtering matrix, which has the full rank  $(N+M)$  under the following assumptions: the  $L$  channels do not share a common zero and  $N \geq (M+1)$ .

The blind identification problem is to find  $\mathcal{H}$  from the sequence,

$$\{\mathbf{r}(k) \quad \text{for} \quad k = 1, 2, \dots, K\}.$$

The subspace method [1] exploits the sample covariance matrix of all channel outputs:  $\mathbf{\Gamma} = E[\mathbf{r}\mathbf{r}^+]$ ,

$\mathbf{\Gamma} = \frac{1}{K} \sum_{k=1}^K \mathbf{r}(k)\mathbf{r}^+(k)$ , where  $K$  is the number of samples and  $^+$  denotes the conjugate transpose. Assume that the signals and the additive noise are independent, stationary and ergodic zero mean complex valued random processes, and as  $K$  becomes large, this matrix has the asymptotical structure:  $\mathbf{\Gamma} = \mathcal{H}\mathbf{\Gamma}_s\mathcal{H}^+ + \mathbf{\Gamma}_n$ , with  $\mathbf{\Gamma}_n = E[\mathbf{n}\mathbf{n}^+]$  the noise covariance matrix and  $\mathbf{\Gamma}_s = E[\mathbf{s}\mathbf{s}^+]$  is the signal covariance matrix.

The goal of blind channel identification and equalization is to identify  $\mathcal{H}$  (channel identification) and to estimate  $\mathbf{s}(k)$  from  $\mathbf{r}(k)$  (channel equalization).

The subspace blind channel identification procedure [1] consists on the estimation of the  $(LN \times 1)$  vector  $\mathbf{h}$  of channel coefficients from the observation vector. Indeed, this approach is based on the eigendecomposition of the data covariance matrix,

$$\mathbf{\Gamma} = [\mathbf{U}_s \quad \mathbf{U}_n] \begin{bmatrix} \mathbf{\Lambda}_s & \\ & \mathbf{\Lambda}_n \end{bmatrix} [\mathbf{U}_s \quad \mathbf{U}_n]^+.$$

The subspace method yields an estimate  $\hat{\mathcal{H}}$  of  $\mathcal{H}$  by solving the equation:  $\mathbf{U}_n^+ \hat{\mathcal{H}} = \mathbf{0}$ , in a least square sense (where  $\hat{\mathcal{H}}$  is subject to the same structure as  $\mathcal{H}$ ). This estimate is uniquely (up to a constant scalar) equal to  $\mathcal{H}$ . From [1], we have:

$$\mathbf{U}_n^+ \mathcal{H} = \mathbf{h}^+ \mathbf{U}_n = \mathbf{0}, \quad (3)$$

with  $\mathbf{U}_n$  is the  $(L(M+1) \times (N+M))$  matrix obtained by stacking the  $L$  filtering matrices  $\mathcal{U}_n^{(i)}$ .

$\mathbf{U}_n = [\mathcal{U}_n^{(0)T} \dots \mathcal{U}_n^{(L-1)T}]^T$ , where,

$$\mathbf{U}_n^{(l)} = \begin{pmatrix} u_1^{(l)} & u_2^{(l)} & \cdots & u_N^{(l)} & \cdots & \cdots & 0 \\ 0 & u_1^{(l)} & u_2^{(l)} & \cdots & u_N^{(l)} & \cdots & 0 \\ \vdots & \cdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & u_1^{(l)} & u_2^{(l)} & \cdots & u_N^{(l)} \end{pmatrix},$$

and  $\mathbf{h} = [\bar{\mathbf{h}}^{(0)}, \dots, \bar{\mathbf{h}}^{(L-1)}]$ , with  $\bar{\mathbf{h}}^{(i)} = [h_0^{(i)}, \dots, h_M^{(i)}]^T$ .

The optimization system derived in [1] is:

$$\hat{\mathbf{h}} = \arg \min_{\|\mathbf{h}\|=1} \mathbf{h}^+ \mathbf{U}_{ss} \mathbf{h}, \quad (4)$$

where,

$$\mathbf{U}_{ss} = \sum_{i=1}^{LN-M-N-1} \mathbf{U}_n^{(i)} \mathbf{U}_n^{(i)+}$$

is the filtering noise projection matrix.

The noise is assumed Gaussian, complex and spatially correlated. Its real and imaginary part are supposed independents, Gaussian with,  $E[\mathbf{n}_i] = \mathbf{0}$ ,  $E[\mathbf{n}_i \mathbf{n}_i^T] = \mathbf{0}$ , and  $E[\mathbf{n}_i \mathbf{n}_i^+] = \mathbf{\Gamma}_n$ .  $\mathbf{\Gamma}_n$  is the noise covariance matrix, the superscripts “\*” and “+” denote conjugate and conjugate transpose, respectively. We consider the noise covariance matrix is band, defined by:

$$\mathbf{\Gamma}_n(i, m) = \begin{cases} 0, & \text{for } |i - m| > K \\ \rho_i, & \text{for } |i - m| \leq K \\ \sigma_i^2, & \text{for } i = m \end{cases} \quad \text{and } i \neq m$$

Where  $\rho_i = \bar{\rho}_i + j\bar{\rho}_i$ ,  $i = 1, \dots, K$ ,  $\rho_i$  are complex variables,  $j^2 = -1$ ,  $\sigma_i^2$  are the noise variance at each receiver, and  $K$  is the spatially noise correlation length.

$$\mathbf{\Gamma}_n = \begin{pmatrix} \sigma_1^2 & \rho_{12} & \cdots & \rho_{1K} & \cdots & 0 \\ \rho_{21}^* & \sigma_2^2 & \rho_{23} & \cdots & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & \rho_{ij}^* & \sigma_i^2 & \cdots & 0 \\ \vdots & \ddots & \cdots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & \rho_{(LN)K}^* & \cdots & \sigma_{LN}^2 \end{pmatrix}.$$

Two manners to give back observation covariance matrix a noise-free matrix: either by subtraction of the noise covariance matrix,  $\mathcal{H}\mathbf{\Gamma}_s\mathcal{H}^+ = \mathbf{\Gamma} - \mathbf{\Gamma}_n$ ; then we have then a “clean” observation covariance matrix; however, we can obtain a negative matrix if  $\mathbf{\Gamma}_n$  is bad-estimated.

Or by whitening; in this case we find again the classical model of communication systems  $(\mathbf{\Gamma}_n^{-\frac{1}{2}} \mathbf{\Gamma} \mathbf{\Gamma}_n^{-\frac{1}{2}})$ . However, this processing is most robust but needs more computational load.

From the data matrix  $\mathbf{\Gamma} = \mathcal{H}\mathbf{\Gamma}_s\mathcal{H}^+ + \mathbf{\Gamma}_n$ , the goal of the first part of this paper is to estimate the noise covariance matrix  $\mathbf{\Gamma}_n$  and in the second part, we estimate,

blindly,  $\mathcal{H}$  from the “clean” obtained matrix  $[\mathcal{H}\mathbf{\Gamma}_s\mathcal{H}^+]$  using the subspace method [1].

### 3. BLIND NOISE ESTIMATION (BNE)

In many applications such as communication systems, it is reasonable to assume the correlation is decreasing along the receivers. That is a widely used model for a colored noise. The correlation rate  $\rho$  is decreasing when the distance between two receivers increases.

In this study, we consider the noise covariance matrix band-Toeplitz with the diagonal values are decreasing, so-called *decreasing band-Toeplitz*. It is the unique assumption to estimate the noise covariance matrix.

The BNE algorithm from the noise covariance matrix estimation is summarized in the following steps:

Step 1: - Estimation and eigendecomposition of the re-

ceived covariance matrix  $\mathbf{\Gamma}$ ;  $\hat{\mathbf{\Gamma}} = \frac{1}{T} \sum_{t=1}^T \mathbf{r}_t \mathbf{r}_t^+$ , with  $T$  is

the number of independent realizations;  $\hat{\mathbf{\Gamma}} = \mathbf{U} \mathbf{\Lambda} \mathbf{U}^+$ , where,  $\mathbf{\Lambda} = \text{diag}[\lambda_1, \dots, \lambda_{LN}]$ , and  $\mathbf{U} = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_{LN}]$ ;  $\lambda_i$  and  $\mathbf{u}_i$  are the eigenvalues and the eigenvectors of the observation covariance matrix, respectively.

- Initialization of the noise covariance matrix:  $\tilde{\mathbf{\Gamma}}_n = \mathbf{0}$ .

Step 2: - Calculation of the matrix:  $\mathbf{W}_{N+M} = \mathbf{U}_S \mathbf{\Lambda}_S^{1/2}$ , with  $\mathbf{U}_S = [\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_{N+M}]$  is the matrix of  $(N+M)$  eigenvectors corresponding to the  $(N+M)$  eigenvalues, and  $\mathbf{\Lambda}_S = \text{diag}[\lambda_1, \dots, \lambda_{N+M}]$  is the matrix of  $(N+M)$  eigenvalues.

- Calculation of the matrix:  $\mathbf{\Delta} = \mathbf{W}_{N+M} \mathbf{W}_{N+M}^+$ .

Step 3: Calculation of:  $\tilde{\mathbf{\Gamma}}_n^{(1)} = K\_band[\hat{\mathbf{\Gamma}} - \mathbf{\Delta}]$ , with  $\tilde{\mathbf{\Gamma}}_n^{(1)}$  is the band noise covariance matrix at first iteration, and  $K\_band[\cdot]$  designates the matrix band with  $(K+1)$  is the bandwidth.

Step 4: Eigendecomposition of the matrix:  $[\hat{\mathbf{\Gamma}} - \tilde{\mathbf{\Gamma}}_n^{(1)}] = \mathbf{V} \mathbf{\Lambda} \mathbf{V}^+$ . The new matrices  $\mathbf{\Delta}$  and  $\tilde{\mathbf{\Gamma}}_n^{(2)}$  are, again, estimated in step 2 and step 3. These iterations are repeated until the improvement of  $\tilde{\mathbf{\Gamma}}_n^{(i)}$ .

Stop test: The algorithm is stopped when the distance between  $\tilde{\mathbf{\Gamma}}_n^{(i)}$  and  $\tilde{\mathbf{\Gamma}}_n^{(i+1)}$  becomes less than some value  $\epsilon$ . We define the distance between  $\tilde{\mathbf{\Gamma}}_n^{(i)}$  and  $\tilde{\mathbf{\Gamma}}_n^{(i+1)}$  as  $\|\tilde{\mathbf{\Gamma}}_n^{(i+1)} - \tilde{\mathbf{\Gamma}}_n^{(i)}\|_F$ , the Frobenius norm of the matrix  $(\tilde{\mathbf{\Gamma}}_n^{(i+1)} - \tilde{\mathbf{\Gamma}}_n^{(i)})$ .

The estimate noise covariance matrix  $\tilde{\mathbf{\Gamma}}_n$  is obtained when the algorithm is stopped.

The matrix  $\tilde{\mathbf{\Gamma}}_n$  is used to “denoise” the received data. In fact, the free-noise received covariance matrix is  $\tilde{\mathbf{\Gamma}} = \hat{\mathbf{\Gamma}} - \tilde{\mathbf{\Gamma}}_n$  or  $\tilde{\mathbf{\Gamma}} = (\tilde{\mathbf{\Gamma}}_n^{-\frac{1}{2}} \hat{\mathbf{\Gamma}} \tilde{\mathbf{\Gamma}}_n^{-\frac{1}{2}})$ . This “clean” matrix

is used to estimate the channel matrix. In order to evaluate its performance, we apply the subspace method [1]. Indeed, Moulines et al. [1], showed that if the subchannels don't share common zeros,  $\mathbf{h}$  is uniquely determined by the noise subspace  $\tilde{\mathbf{U}}_n$ , the subspace estimator is given by:

$$\hat{\mathbf{h}} = \arg \min_{\|\mathbf{h}\|=1} \mathbf{h}^+ \hat{\mathbf{U}}_{ss} \mathbf{h}, \text{ where } \hat{\mathbf{U}}_{ss} \text{ is the filtering noise}$$

projection matrix estimated from the "clean" data covariance matrix. This estimator does not require the knowledge of the source covariance as long as  $\Gamma_s > 0$ . We also compare our result to the modified subspace (MSS) method [4].

#### 4. PERFORMANCE EVALUATION

To demonstrate the efficiency of the proposed algorithm, some computer simulations have been conducted. In the following simulations, we take the parameters described in [1], in fact the number of virtual channels is  $L = 4$ ; the width of the temporal window is  $N = 10$ ; the degree of the ISI is  $M = 4$ , the channel coefficients are given by [1]:

$h_0$	$h_1$	$h_2$	$h_3$
-0.049+0.359j	0.443-0.0364j	-0.211-0.322j	0.417-0.030j
0.482+0.569j	1	-0.199-0.918j	1
-0.556+0.587j	0.921-0.194j	1	0.873-0.145j
1	0.189-0.208j	-0.284-0.524j	0.285+0.309j
-0.171+0.061j	-0.087-0.054j	0.136-0.190j	-0.049+0.161j

Table 1: Four virtual complex channels.

for all these simulations, the number of data samples used to estimate each  $\mathbf{h}$  ranges from 100 to 1000 in steps of 100.

The root mean-square error (*RMSE*) defined, below, is employed as a performance measure of the input estimates:

$$RMSE = \frac{1}{\|\mathbf{H}\|} \sqrt{\frac{1}{K} \sum_{i=1}^K \|\mathbf{H}_i - \mathbf{H}\|^2}, \text{ where } K \text{ is the number of trials (100 in our cases) and } \mathbf{H}_i \text{ is the estimate of the inputs from the } i\text{th} \text{ trial.}$$

The signal to noise ratio (*SNR*) is defined as:

$$SNR = 10 \log_{10} \frac{E\{\|\mathbf{H}\mathbf{s}(k)\|^2\}}{E\{\|\mathbf{n}(k)\|^2\}}. \text{ We define the Frobenius norm of estimation error (EE) of the noise covariance matrix as: } EE = \|\Gamma - (\mathbf{H}\Gamma_s\mathbf{H}^+ + \Gamma_n)\|_F.$$

We compare the presented algorithm with the existing methods such as the modified subspace approach (MSS) [4]. This comparison is based on the root mean square error of the channel matrix estimates. We recall, this approach in the following: Let  $\Gamma(\tau) = \mathcal{H}\mathbf{J}(\tau)\mathcal{H}^+ + \Gamma_n(\tau)$ , where  $\mathbf{J}(\tau)$  is the  $(N+M) \times (N+M)$  shift matrix. In [4], one assumes that  $\Gamma_n(\tau) = \mathbf{0}$  as long as  $\tau \geq N$ . Therefore, we have the relation  $\Gamma(\tau) = \mathcal{H}\mathbf{J}(\tau)\mathcal{H}^+$  for  $\tau \geq N$ . At the time lag  $\tau = N$ ,  $\Gamma(N) =$

$\mathcal{H}(\mathbf{J}(N) + \mathbf{J}(N)^+)\mathcal{H}^+$ , the matrix  $\Gamma(N)$  is used to estimate the channel parameters.

The Figures (1a and 1b) present the root square-mean error (RMSE) of the parameters estimates for a band-Toeplitz noise covariance matrix and the Frobenius norm of estimation of error (EE) of the noise covariance matrix versus number of samples.

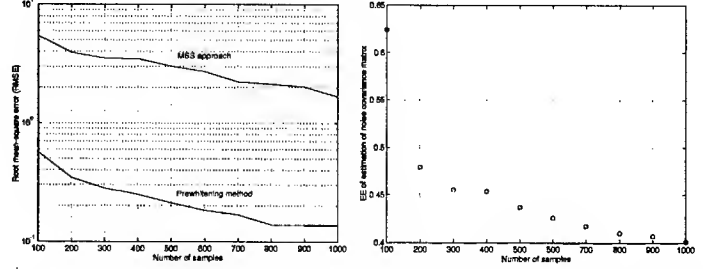


Figure 1: (a) Root square-mean error (RMSE) of the parameters estimates (band-Toeplitz noise covariance matrix). (b) Frobenius norm of estimation of error (EE) of the noise covariance matrix (band-Toeplitz noise covariance matrix) versus number of samples

In the case of a band noise covariance matrix with a correlation length  $K = 4$ , we have Figures (2a and 2b), versus *SNR* between 0 dB to 16 dB.

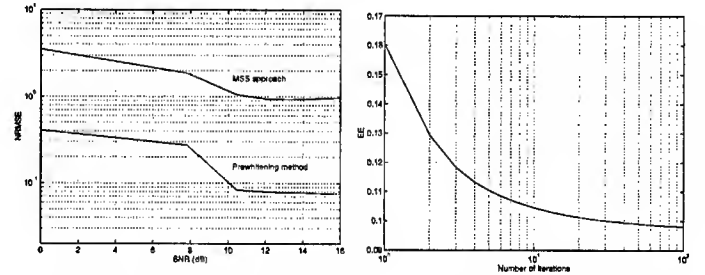


Figure 2: (a) Root mean-square error of the parameters estimates (band-Toeplitz noise covariance matrix ( $K = 4$ )) versus *SNR*. (b) Frobenius norm of the estimation of error (EE) of noise covariance matrix as a function of number of iterations.

We study, the influence of the correlation length versus the error of the noise covariance matrix estimation Figure (3a) and the channel parameters Figure (3b). In fact, the correlation length varies between  $K = 1$  and  $K = 4$ , with *SNR* = 3 dB.

The normalized error (*NE*) is defined by,  $NE = \frac{\|\mathbf{H}_i - \mathbf{H}\|}{\|\mathbf{H}\|}$ .

We consider the noise covariance matrix band, and we estimate the normalized error and the Frobenius norm versus of different scenarios of the channel matrix (Figures (4a and 4b)).

These simulations show that the processing which consists to first estimation of the noise covariance matrix and prewhitening the observation has many advantages, is more efficient then the modified subspace (MSS) approach [4]. The use of the denoised subspace

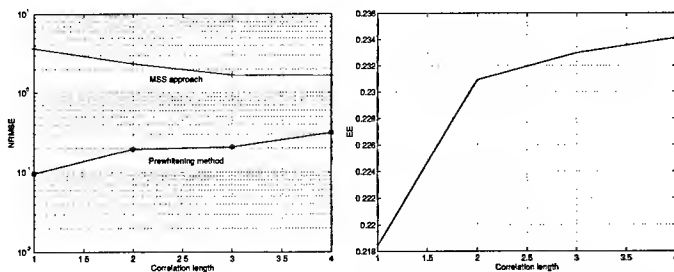


Figure 3: (a) Root mean-square error of the parameters estimates versus correlation length. (b) Frobenius norm of the estimation of error (EE) of noise covariance matrix as a function of correlation length.

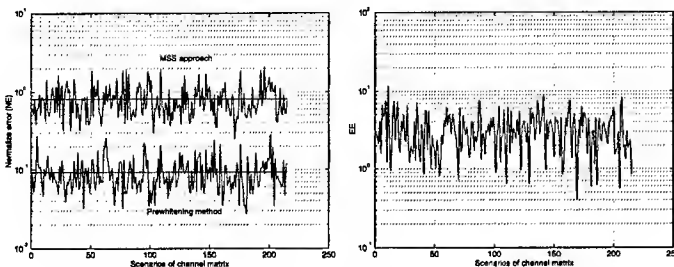


Figure 4: (a) Normalized error (NE) of the parameters estimates versus scenarios of channel matrix when the noise covariance matrix is band. (b) Frobenius norm of the estimation of error (EE) of band noise covariance matrix as a function of scenarios of channel matrix.

method presented in this paper becomes interesting in the case of low  $SNR$  and when the noise covariance matrix is band. When the length correlation increases, the interest of the estimation of the noise increases also. Several computer simulations confirm these conclusions.

This algorithm can be, also, applied, naturally, for other blind channel identification methods such as XBM, TXK ...[2, 3] disregard of the system type used.

## 5. CONCLUSION

To estimate, blindly, the noise than the channel parameters, an algorithm was presented. We have considered a spatially correlated noise, with only the assumption that the matrix noise is band-Toeplitz, than by an iterative algorithm using the eigenstructure, we have estimated the noise parameters. In order to use a "clean" data for the the estimation of the channel matrix, the estimated noise matrix was used for "prewhitening" the observations. The subspace approach was, then, applied for the blind estimation of the channel parameters.

## REFERENCES

- [1] E. Moulines, P. Duhamel, J.F. Carodoso, and S. Mayrargue, "Subspace methods for the blind identification of multichannel fir filters," *IEEE Trans. on Signal Processing*, vol. 43, no. 2, pp. 516-525, Feb. 1995.
- [2] L. Tong, G. Xu, and T. Kailath, "Blind identification and equalization based on second-order statistics: A time domain approach," *IEEE Trans. on information Theory*, vol. 40, no. 2, pp. 340-349, Mar. 1994.
- [3] J. Xavier, V. Barroso, and J. Moura, "Closed-form blind channel identification and source separation in sdma systems through correlative coding," *accepted for IEEE Journal on Selected Areas on Communication, Special Issue on Signal Processing for Wireless Communications*, 1997.
- [4] K. Abed-Meraim, Y. Hua, P. Loubaton, and E. Moulines, "Subspace method for blind identification of multichannel fir systems in noise field with unknown spatial covariance," *IEEE Signal Processing Letters*, vol. 4, no. 5, pp. 135-137, May 1997.
- [5] J. G. Proakis, "Digital communication.", 3rd ed. Mc Graw-Hill, 1995.
- [6] J. Böhme and D. Krauss, "On least squares methods for direction of arrival estimation in the presence of unknown noise fields," in *Proceedings IEEE-ICASSP'88*, New York, NY, Apr. 1988, pp. 2833-2836.
- [7] B. Friedlander and A. J. Weiss, "Direction finding using noise covariance modeling," *IEEE Trans. on Signal Processing*, vol. SP-43, no. 7, pp. 1557-1567, Jul. 1995.
- [8] A. Paulraj and T. Kailath, "Eigenstructure methods for direction of arrival estimation in the presence of unknown noise field," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 34, no. 1, pp. 276-280, Feb. 1986.
- [9] P. Stoica, M. Viberg, and B. Ottersten, "Instrumental Variable approach to array processing in spatially correlated noise fields," *IEEE Trans. on Signal Processing*, vol. 42, no. 1, pp. 121-133, 1994.
- [10] K. M. Wong, J. Reilly, Q. Wu, and S. Qiao, "Estimation of the direction-of-arrival of signals in the unknown correlated noise, part I: The MAP approach and its implementation," *IEEE Trans. on Signal Processing*, vol. 40, no. 8, pp. 2007-2017, Aug. 1992.
- [11] J.-P. Le Cadre, "Parametric methods for spatial signal processing in the presence of unknown colored noise fields," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-37, no. 7, pp. 965-983, Jul. 1989.
- [12] M. Wax, "Detection and localization of multiple sources in noise with unknown covariance," *IEEE Trans. on Signal Processing*, vol. 40, no. 1, pp. 245-249, Sep. 1991.
- [13] V. Barroso, J. Moura, and J. Xavier, "Blind array channel division multiple access (achdma) for mobile communications," *IEEE Trans. on Signal Processing*, vol. 46, no. 3, pp. 516-525, Mar. 1998.

# MULTIUSER DETECTION IN IMPULSIVE NOISE VIA SLOWEST DESCENT SEARCH

Predrag Spasojević

WINLAB,  
Dept. of Electrical and Computer Eng.,  
Rutgers University,  
Piscataway, NJ 08854.

Xiaodong Wang

Department of Electrical Engineering,  
Texas A&M University,  
College Station, TX 77843.

## ABSTRACT

A new technique is proposed for robust multiuser detection in the presence of non-Gaussian ambient noise. This method is based on minimizing a certain cost function (e.g., the Huber penalty function) over a discrete set of candidate user bit vectors. The set of candidate points are chosen based on the so-called "slowest-descent search", starting from the estimate closest to the unconstrained minimizer of the cost function, and along mutually orthogonal directions where this cost function grows the slowest. The extension of the proposed technique to multi-user detection in unknown multi-path fading channels is also proposed. Simulation results show that this new technique offers substantial performance improvement over the recently proposed robust multiuser detectors, with little attendant increase in computational complexity.

## 1. INTRODUCTION

Recently, a robust multiuser detection technique is developed in [4] for demodulating multiuser signals in the presence of both multiple-access interference and impulsive ambient channel noise. This technique is based on the  $M$ -estimation method for robust regression, and is essentially the robustized version of the linear decorrelating multiuser detector. Although this robust multiuser detector offers significant performance gain over the linear decorrelator in impulsive noise, there is still a large gap between its performance and that of the maximum likelihood (ML) multiuser detector. However, the computational complexity of the ML detection is quite high, and moreover, the ML detection requires the knowledge of the exact probability distribution of the noise, which may not be available to the receiver. Hence, it is of interest to develop robust, low-complexity, and near-optimal multiuser detection techniques for non-Gaussian noise channels. Furthermore, it is of high importance to have the ability of successfully extending this method to more general asynchronous unknown multi-path fading channels. Described issues are subjects of this paper.

P. Spasojević was supported in part by the WINLAB/Lucent Technologies Wireless Post-Doctoral Fellowship. X. Wang was supported in part by the NSF grant CAREER CCR-9875314.

## 2. SYNCHRONOUS SYSTEM MODEL

First consider the following discrete-time synchronous CDMA signal model. At any time instant, the received signal is the superposition of  $K$ -user signals, plus the ambient noise, given by

$$\mathbf{r} = \sum_{k=1}^K \alpha_k b_k \mathbf{s}_k + \mathbf{n} = \mathbf{S} \mathbf{A} \mathbf{b} + \mathbf{n}, \quad (1)$$

where  $\mathbf{s}_k = \frac{1}{\sqrt{N}}[s_{k1} \cdots s_{kN}]^T$  is the normalized signature sequence of the  $k$ -th user;  $N$  is the processing gain;  $b_k \in \{+1, -1\}$  and  $\alpha_k$  are respectively the data bit and the complex amplitude of the  $k$ -th user;  $\mathbf{S} \triangleq [\mathbf{s}_1 \cdots \mathbf{s}_K]$ ;  $\mathbf{A} \triangleq \text{diag}(\alpha_1, \dots, \alpha_K)$ ;  $\mathbf{b} \triangleq [b_1 \cdots b_K]^T$ ; and  $\mathbf{n} = [n_1 \cdots n_N]^T$  is a vector of independent and identically distributed (i.i.d.) ambient noise samples with independent real and imaginary components. Denote

$$\mathbf{y} \triangleq \begin{bmatrix} \Re\{\mathbf{r}\} \\ \Im\{\mathbf{r}\} \end{bmatrix}, \quad \mathbf{\Psi} \triangleq \begin{bmatrix} \mathbf{S}\Re\{\mathbf{A}\} \\ \mathbf{S}\Im\{\mathbf{A}\} \end{bmatrix}, \quad \mathbf{v} \triangleq \begin{bmatrix} \Re\{\mathbf{n}\} \\ \Im\{\mathbf{n}\} \end{bmatrix},$$

where  $\mathbf{v}$  is a real noise vector consisting of  $2N$  i.i.d. samples. Then (1) can be written as

$$\mathbf{y} = \mathbf{\Psi} \mathbf{b} + \mathbf{v}. \quad (2)$$

It is assumed that each element  $v_j$  of  $\mathbf{v}$  follows a two-term Gaussian mixture distribution, i.e.,

$$v_j \sim (1 - \epsilon) \mathcal{N}(0, \nu^2) + \epsilon \mathcal{N}(0, \kappa \nu^2), \quad (3)$$

with  $0 < \epsilon < 1$  and  $\kappa > 1$ . Here the term  $\mathcal{N}(0, \nu^2)$  represents the nominal ambient noise, and the term  $\mathcal{N}(0, \kappa \nu^2)$  represents an impulsive component. The probability that impulses occur is  $\epsilon$ . Note that the overall variance of the noise sample  $v_j$  is

$$\frac{\sigma^2}{2} \triangleq (1 - \epsilon) \nu^2 + \epsilon \kappa \nu^2. \quad (4)$$

We have  $\text{Cov}\{\mathbf{v}\} = \frac{\sigma^2}{2} \mathbf{I}_{2N}$ ; and  $\text{Cov}\{\mathbf{n}\} = \sigma^2 \mathbf{I}_N$ . The model (3) serves as an approximation to the more fundamental Middleton Class A noise model [2, 5], and has been

used extensively to model physical noise arising in radio and acoustic channels. Recently, it has been shown that another class of non-Gaussian distributions, the  $\alpha$ -stable distributions, can be well approximated by a finite mixture of Gaussians [1]. In what follows, we consider the problem of detecting the transmitted symbols  $\mathbf{b}$  of all users based on the signal model (2).

### 3. EXHAUSTIVE-SEARCH DETECTION AND DECORRELATIVE DETECTION

In this section, we give a unified description of a number of approaches to the problem of multiuser detection in non-Gaussian noise. There are primarily two categories of such detectors for estimating  $\mathbf{b}$  from  $\mathbf{y}$  in (2), all based on minimizing the sum of a certain function  $\rho$  of the chip residuals

$$\mathcal{C}(\mathbf{b}; \mathbf{y}) \triangleq \sum_{j=1}^{2N} \rho(y_j - \xi_j^T \mathbf{b}), \quad (5)$$

where  $\xi_j^T$  is the  $j$ -th row of the matrix  $\Psi$ .

- Exhaustive-search detector:

$$\mathbf{b}^e = \arg \min_{\mathbf{b} \in \{+1, -1\}^K} \mathcal{C}(\mathbf{b}; \mathbf{y}). \quad (6)$$

- Decorrelative detector:

$$\boldsymbol{\beta} = \arg \min_{\mathbf{b} \in \mathbb{R}^K} \mathcal{C}(\mathbf{b}; \mathbf{y}), \quad (7)$$

$$\mathbf{b}^* = \text{sign}(\boldsymbol{\beta}). \quad (8)$$

It is seen that the exhaustive-search detection is based on the discrete minimization of the cost function  $\mathcal{C}(\mathbf{b}; \mathbf{y})$ , over  $2^K$  candidate points; whereas the decorrelative detection is based on the continuous minimization of the same cost function. In general, the optimization problem (7) can be solved iteratively according to the following steps [4]

$$\mathbf{z}^l = \Psi(\mathbf{y} - \Psi \boldsymbol{\beta}^l), \quad (9)$$

$$\boldsymbol{\beta}^{l+1} = \boldsymbol{\beta}^l + (\Psi^T \Psi)^{-1} \Psi^T \mathbf{z}^l, \quad l = 0, 1, \dots, \quad (10)$$

We consider the following three choices of the penalty function  $\rho(\cdot)$  in (5), corresponding to different forms of detectors:

- Log-likelihood penalty function:

$$\rho_{\text{ML}}(x) \triangleq -\log f(x), \quad \psi_{\text{ML}}(x) = -\frac{f'(x)}{f(x)}, \quad (11)$$

where  $f(\cdot)$  denotes the probability density function (pdf) of the noise sample. In this case, the exhaustive-search detector (6) corresponds to the ML detector; and the decorrelative detector (8) corresponds to the ML decorrelator [4].

- Least-square penalty function:

$$\rho_{\text{LS}}(x) \triangleq \frac{1}{2}x^2, \quad \psi_{\text{LS}}(x) = x. \quad (12)$$

In this case, the exhaustive-search detector (6) corresponds to the ML detector based on a Gaussian noise assumption; and the decorrelative detector (8) corresponds to the linear decorrelator.

- Huber penalty function:

$$\rho_{\text{H}}(x) = \begin{cases} \frac{x^2}{\sigma^2}, & \text{if } |x| \leq \frac{c\sigma^2}{2}, \\ c|x| - \frac{c^2\sigma^2}{4}, & \text{if } |x| > \frac{c\sigma^2}{2}, \end{cases} \quad (13)$$

$$\psi_{\text{H}}(x) = \begin{cases} \frac{x}{\sigma^2}, & \text{if } |x| \leq \frac{c\sigma^2}{2}, \\ c \text{sign}(x), & \text{if } |x| > \frac{c\sigma^2}{2}. \end{cases} \quad (14)$$

where  $\frac{\sigma^2}{2}$  is the noise variance given by (4), and  $c = \frac{3}{2\sigma}$  is a constant. In this case, the exhaustive-search detector (6) corresponds to the discrete minimizer of the Huber cost function; and the decorrelative detector (8) corresponds to the robust decorrelator proposed in [4].

### 4. SLOWEST-DESCENT-SEARCH DETECTION

Clearly the optimal performance is achieved by the exhaustive search detector with the log-likelihood penalty function, i.e., the ML detector. As will be seen in Section 5, the performance of the exhaustive search detector with the Huber penalty function is close to that of the ML detector, while this detector does not require the knowledge of the exact noise pdf. However computational complexity of the exhaustive search detector (6) is on the order of  $O(2^K)$ . We next propose a local search approach to approximating the solution to (6). The basic idea is to minimize the cost function  $\mathcal{C}(\mathbf{b}; \mathbf{y})$  over a subset  $\Omega$  of the discrete parameter set  $\{-1, +1\}^K$  that is close to the continuous stationary point  $\boldsymbol{\beta}$  given by (7). More precisely, we approximate the solution to (6) by

$$\mathbf{b}^s \triangleq \arg \min_{\mathbf{b} \in \Omega} \mathcal{C}(\mathbf{b}; \mathbf{y}). \quad (15)$$

In the *slowest descent* method [3], the candidate set  $\Omega$  consists of the discrete parameters chosen such that they are in the neighborhood of  $\mathbf{Q}$  ( $Q \leq K$ ) lines in  $\mathbb{R}^K$ , which are defined by the stationary point  $\boldsymbol{\beta}$  and the  $Q$  eigenvectors of the Hessian matrix  $\nabla_{\mathbf{b}}^2 \mathcal{C}(\boldsymbol{\beta})$  of  $\mathcal{C}(\mathbf{b}; \mathbf{y})$  at  $\boldsymbol{\beta}$  corresponding to the  $Q$  smallest eigenvalues. The basic idea of this method is explained next.

*Slowest-Descent Search:* The basic idea of the slowest-descent search method is to choose the candidate points in  $\Omega$  such that they are closest to a line  $(\boldsymbol{\beta} + \mu \mathbf{g})$  in  $\mathbb{R}^K$ , originating from  $\boldsymbol{\beta}$  and along a direction  $\mathbf{g}$ , where the cost function  $\mathcal{C}(\mathbf{b}; \mathbf{y})$  increases at the slowest rate. Given any line in  $\mathbb{R}^K$ , there are at most  $K$  points where the line intersects the coordinate hyper-planes (e.g.,  $\boldsymbol{\beta}^1$  and  $\boldsymbol{\beta}^2$  in Figure 1 for  $K = 2$ ). The set of intersection points corresponding to a line defined by  $\boldsymbol{\beta}$  and  $\mathbf{g}$  can be expressed as

$$\{\boldsymbol{\beta}^i = \boldsymbol{\beta} - \mu_i \mathbf{g} : \mu_i = \beta_i / g_i\}_{i=1}^K, \quad (16)$$

where  $\beta_i$  and  $g_i$  denote the  $i$ -th elements of the respective vectors  $\boldsymbol{\beta}$  and  $\mathbf{g}$ . Each intersection point  $\boldsymbol{\beta}^i$  has only its  $i$ -th component equal to zero, i.e.,  $\beta_i^i = 0$ .

Any point on the line except for an intersection point has a unique closest candidate point in  $\{+1, -1\}^K$ . An intersection point is of equal distance from its two neighboring candidate points, e.g.,  $\boldsymbol{\beta}^1$  is equi-distant to  $\mathbf{b}^1$  and  $\mathbf{b}^2$  in Figure 1(a). Two neighboring intersection points share

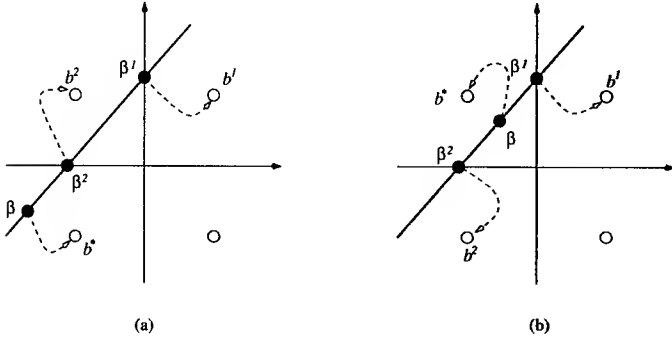


Figure 1: One-to-one mapping from  $\{\beta, \beta^1, \dots, \beta^K\}$  to  $\Omega \triangleq \{b^*, b^1, \dots, b^K\}$  for  $K = 2$ . Each intersection point  $\beta^i$  is of equal distance from its two neighboring candidate points.  $b^i$  is chosen to be one of these two candidate points that is on the opposite side of the  $i$ -th coordinate hyper-plane with respect to  $b^*$ .

a unique closest candidate point, e.g.,  $\beta^1$  and  $\beta^2$  share the nearest candidate point  $b^2$  in Figure 1(a). Note that  $b^*$  in (8) is the candidate point closest to  $\beta$ . By carefully selecting one of the two candidate points closest to each intersection point to avoid choosing the same point twice, one can specify  $K$  distinct candidate points in  $\{+1, -1\}^K$  that are closest to the line  $(\beta + \mu g)$ . To that end, consider the following set

$$\{b^i \in \{-1, +1\}^K : b_k^i = \begin{cases} \text{sign}(\beta_k^i), & k \neq i \\ -b_k^*, & k = i \end{cases}\}_{i=1}^K. \quad (17)$$

It is seen that (17) assigns to each intersection point  $\beta^i$  the closest candidate point  $b^i$  that is on the opposite side of the  $i$ -th coordinate hyper-plane from  $b^d$  [cf. Figure 1 (a) (b)].

In general, the slowest-descent search method chooses the candidate set  $\Omega$  in (15) as follows:

$$\begin{aligned} \Omega &= \{b^d\} \cup \bigcup_{q=1}^Q \{b^{q,\mu} \in \{-1, +1\}^K : \\ b_k^{q,\mu} &= \begin{cases} \text{sign}(\beta_k - \mu g_k^q), & \text{if } \beta_k - \mu g_k^q \neq 0 \\ -b_k^*, & \text{if } \beta_k - \mu g_k^q = 0 \end{cases}, \\ g^q &\text{ is the } q\text{-th smallest eigenvector of } \nabla_C^2, \\ \mu &\in \left\{ \frac{\beta_1}{g_1^q}, \dots, \frac{\beta_K}{g_K^q} \right\} \}. \end{aligned} \quad (18)$$

Hence,  $\{b^{q,\mu}\}_\mu$  contains the  $K$  closest neighbors of  $\beta$  in  $\{-1, +1\}^K$  along the direction of  $g^q$ . Note that  $\{g^q\}_{q=1}^Q$  represent the  $Q$  mutually orthogonal directions where the cost function  $C(b; y)$  grows the slowest from the minimum point  $\beta$ . (In case of the log-likelihood penalty function, this corresponds to the situation where the likelihood function drops the slowest from its peak, hence the name "slowest descent") Intuitively, the solution to (6) is most likely found in this neighborhood.

For the three types of the penalty functions, the Hessian matrix at the stationary points are given respectively by

$$\rho_{ML} : \quad \nabla_C^2(\beta) = \Psi^T \text{diag} \left\{ \rho_{ML}''(y_j - \xi_j^T \beta) \right\} \Psi, \quad (19)$$

$$\rho_{LS} : \quad \nabla_C^2(\beta) = \Psi^T \Psi, \quad (20)$$

$$\rho_H : \quad \nabla_C^2(\beta) = \Psi^T \text{diag} \left\{ \delta \left( |y_j - \xi_j^T \beta| \leq \frac{c\sigma^2}{2} \right) \right\} \Psi, \quad (21)$$

where in (19)  $\rho_{ML}''(x) = \psi_{ML}^2(x) - f''(x)/f(x)$  and in (21) the indicator function  $\delta(y \leq a) = 1$  if  $y \leq a$  and 0 otherwise; hence in this case those rows of  $\Psi$  with large residual signals as a possible result of impulsive noise are nullified, whereas other rows of  $\Psi$  are not affected.

Finally we summarize the slowest-descent search algorithm for multiuser detection in non-Gaussian noise. Given a penalty function  $\rho(\cdot)$ , this algorithm solves the discrete optimization problem (15) according to the following steps:

1. Compute the continuous stationary point  $\beta$  in (7) using the iteration (9)-(10);
2. Compute the Hessian matrix  $\nabla_C^2(\beta)$  given by (19) or (20) or (21), and its  $Q$  smallest eigenvectors  $g^1, \dots, g^Q$ ;
3. Solve the discrete optimization problem defined by (15) and (18) by an exhaustive search (over  $(KQ+1)$  points).

## 5. EXTENSION TO AN UNKNOWN MULTIPATH CHANNEL

In this section, we extend the slowest descent multiuser detection techniques developed above to the asynchronous CDMA system with multipath distortion. Following [4], [7], and references therein,  $r[i]$ , the vector consisting of a number of stacked one-symbol length vectors that affect the current symbol interval  $i$  can be expressed as follows:

$$r[i] = H b[i] + n[i]. \quad (22)$$

Here,  $b[i]$  and  $n[i]$  are stacked symbol vectors, and  $H$  is the unknown channel matrix.

We can rewrite (22) as

$$r[i] = H \theta[i] + n[i] = U_s \zeta[i] + n[i].$$

Here orthonormal column vectors of  $U_s$  span the column space of  $H$  and can be obtained using an eigen-decomposition of the received signal autocorrelation matrix (see [6]). The estimation of the channel matrix  $H$  is based on the users' signature sequences and the noise subspace estimated from the auto-correlation eigen-decomposition (see [7]).

We next obtain the robust estimate of  $\zeta[i]$  based on the complex version of the decorrelative iterations (9)-(10) for an (e.g., Huber) objective function

$$z^l = \psi(r - U_s \zeta^l), \quad (23)$$

$$\zeta^{l+1} = \zeta^l + U_s^H z^l, \quad l = 0, 1, 2, \dots \quad (24)$$



where  $H$  is the Hermitian operator.  $\theta[i]$  can be estimated as follows

$$\theta[i] = (H^H H)^{-1} H^H U_s \zeta[i].$$

Note that

$$H\theta[i] = H_0 A b[i] + H_{\#} \theta_{\#}[i], \quad (25)$$

where the term  $H_0 A b[i]$  contains the signal carrying the current bits  $b[i]$ ; and the term  $H_{\#} \theta_{\#}[i]$  contains the signal carrying the previous and future bits  $\{b[l]\}_{l \neq i}$ , i.e., intersymbol interference.  $A$  holds unknown phases which are estimated separately from the channel as demonstrated below. We subtract the estimated intersymbol interference from  $r[i]$  to obtain

$$\tilde{r}[i] \triangleq r[i] - H_{\#} \theta_{\#}[i] \quad (26)$$

$$= H_0 A b[i] + n[i], \quad (27)$$

We can now set

$$\Psi \triangleq \begin{bmatrix} H_0 \Re\{A\} \\ H_0 \Im\{A\} \end{bmatrix},$$

and use the methods described in previous sections to derive decorrelative and the slowest-descent estimates of  $b[i]$  based on  $\tilde{r}[i]$ .

### Estimation of $A$

We next consider the estimation of the complex amplitudes

**A.** Following (25), we have [recall that  $A \triangleq \text{diag}(\alpha_1, \dots, \alpha_K)$ ].

$$\theta_k = \alpha_k b_k + \tilde{n}_k, \quad k = 1, \dots, K. \quad (28)$$

Since  $b_k \in \{-1, +1\}$ , it follows from (28) that  $\theta_k$  form two clusters centered at respectively  $\alpha_k$  and  $-\alpha_k$ . Let  $\alpha_k = \rho_k e^{j\phi_k}$ , a simple estimator of  $\alpha_k$  is given by  $\hat{\alpha}_k = \hat{\rho}_k e^{j\hat{\phi}_k}$  with

$$\begin{aligned} \hat{\rho}_k &= E\{|\theta_k|\}, \\ \hat{\phi}_k &= \end{aligned}$$

$$\begin{cases} E\{\angle[\theta_k \text{sign}(\Re\{\theta_k\})]\}, & \text{if } E\{|\Re\{\theta_k\}|\} > E\{|\Im\{\theta_k\}|\} \\ E\{\angle[\theta_k \text{sign}(\Im\{\theta_k\})]\}, & \text{if } E\{|\Re\{\theta_k\}|\} < E\{|\Im\{\theta_k\}|\} \end{cases},$$

where the operator  $E(\cdot)$  denotes sample average. Note that the above estimate of the phase  $\phi_k$  has an ambiguity of  $\pi$ , which necessitates differential encoding and decoding of data.

## 6. SIMULATION RESULTS

For simulations, we assume a synchronous CDMA system with a processing gain  $N = 15$ , number of users  $K = 6$ , no phase offset and equal amplitudes of user signals, i.e.,  $\alpha_k = 1$ ,  $k = 1, \dots, K$ . User 1 signature  $s_1$  sequence is generated randomly and kept fixed throughout simulations. Signature sequences of Users 2 through  $K$  are generated by a circularly shifting the sequence of User 1.

For each of the three penalty functions Figure 2 presents the symbol error performance of the decorrelative detector,

the slowest descent detector with 2 search directions, and the exhaustive detector. Searching further slowest descent directions does not improve the performance in this case. We observe that for all three criteria the performance of the slowest descent detector is close to the performance of its respective exhaustive maximization version. All detectors are significantly better than the LS based detectors.

For the multi-path channel case the following is assumed: processing gain  $N = 15$ , number of users  $K = 6$  each user's channel has 3 paths and a delay spread of up to one symbol interval. The complex gains, the delays of each user's channel, and user signature sequences are generated randomly. The chip pulse is a raised cosine pulse with roll-off factor 0.5. The path gains are normalized so that each user's signal arrives at the receiver with unit energy. The over-sampling factor is 2 and the number of stacked vectors in (22) (the smoothing factor) is 2.

Figure 3 demonstrates the performance of the Huber-based slowest-descent method with one and two search directions, the decorrelative Huber detector, and the blind decorrelator from [6]. Most of the performance gain offered by the slowest-descent method is obtained by searching along only one direction. Over 1 dB of gain is obtained relative to the the decorrelative estimate. The blind approach [6] performs poorly for this system.

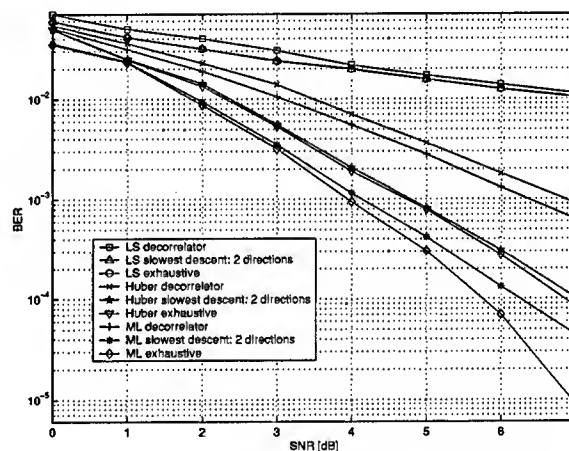


Figure 2: Symbol error performance of a synchronous DS-CDMA system with  $N = 15$ ,  $K = 8$ ,  $\epsilon = 0.01$ ,  $\kappa = 100$ .

## 7. CONCLUSION

We have developed a new robust multiuser detection technique based on the method of slowest-descent search. By searching only over one or two directions, this method offers significant performance improvement over the recently proposed robust decorrelating detector in impulsive noise. The proposed approach has been extended to multi-path fading channels where complex channels and signal phases of all users have to be estimated blindly.

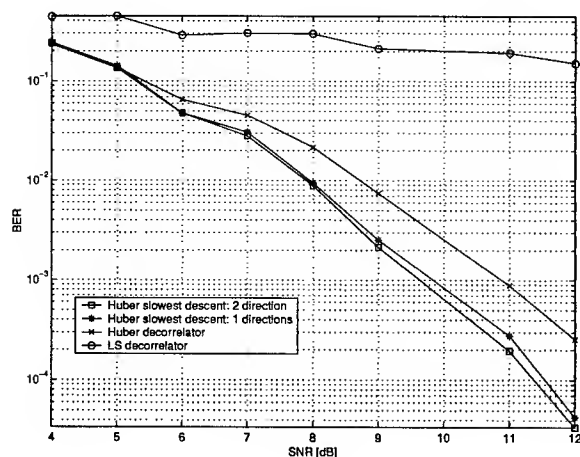


Figure 3: Symbol error performance of an asynchronous DS-SS-CDMA system with  $N = 15$ ,  $K = 8$ ,  $\epsilon = 0.01$ ,  $\kappa = 100$ , in an unknown multi-path channel with 3 randomly generated path coefficients per user.

## REFERENCES

- [1] E.E. Kuruoglu and C. Molina and W.J. Fitzgerald. Approximation of  $\alpha$ -stable probability densities using finite mixtures of Gaussians. In *Proc. EUSIPCO'98*, Rhodes, Greece, September 1998.
- [2] D. Middleton. Non-gaussian noise models in signal processing for telecommunications: New methods and results for class A and class B noise models. *IEEE Trans. Inform. Theory*, 45(4):1122–1129, May 1999.
- [3] P. Spasojević. Sequence and channel estimation for channels with memory. Department of Electrical Engineering, Texas A&M University 1999.
- [4] X. Wang and H.V. Poor. Robust multiuser detection in non-Gaussian channels. *IEEE Trans. Sig. Proc.*, 47(2):289–305, Feb. 1999.
- [5] S.M. Zabin and H.V. Poor. Efficient estimation of the class A parameters via the EM algorithm. *IEEE Trans. Inform. Theory*, 37(1):60–72, Jan. 1991.
- [6] X. Wang and H.V. Poor. Blind multiuser detection: A subspace approach. *IEEE Trans. Inform. Theory*, 44(2):677–691, Mar. 1998.
- [7] P. Spasojević, X. Wang, and A. Høst-Madsen, "Nonlinear group-blind multiuser detection," *Technical Report, WINLAB, Rutgers Univ.*, July 2000.

# MAXIMUM LIKELIHOOD DELAY-DOPPLER IMAGING OF FADING MOBILE COMMUNICATION CHANNELS

Linda M. Davis<sup>†</sup>   Iain B. Collings<sup>‡</sup>   Robin J. Evans<sup>\*</sup>

<sup>†</sup>Global Wireless Systems Research  
Bell Laboratories  
Lucent Technologies, AUSTRALIA  
lindadavis@lucent.com

<sup>‡</sup>School of Electrical & Information Engineering  
University of Sydney, AUSTRALIA

<sup>\*</sup>Dept. Electrical & Electronic Engineering  
University of Melbourne, AUSTRALIA

## ABSTRACT

This paper presents a new recursive algorithm for maximum likelihood estimation of the delay-Doppler characteristics of fast-fading mobile communication channels. The channel is modelled as an FIR filter with rapidly varying complex coefficients. The parameters of interest are the mean channel taps and the tap covariance. The structure of the channel tap covariance matrix is exploited to provide convergence to constrained channel estimates.

## 1. INTRODUCTION

Maximum likelihood constrained covariance estimation for directly observable processes in additive noise has received considerable attention [1, 2, 3, 4] since many algorithms in spectral analysis rely on knowledge of the covariance matrix. Applications include harmonic retrieval, beamforming and direction of arrival estimation. In many such cases, the system of interest is shift-invariant and the true covariance matrix is known to be Hermitian Toeplitz as well as positive semidefinite. This structure may be used in obtaining realistic covariance matrix estimates, and in addition may be exploited in to provide fast convergence to constrained estimates and aid subsequent processing (e.g. inverses, eigendecomposition etc.).

In this paper, we consider the extension of constrained covariance estimation to the case where the process of interest is observed through convolution with a known signal in addition to the additive noise. This problem arises in delay-Doppler radar imaging [5] and delay-Doppler imaging of fast-fading mobile communication channels [6]. In these situations, the underlying reflectance process has a time-varying impulse response,  $f_{k,\epsilon}$ , and therefore is two-dimensional (in time,  $k$ , and

delay,  $\epsilon$ ). The delay-Doppler image of a reflectance process is also known as the scattering function, and is related to the covariance matrix by a Fourier transform (in the time axis indexed by  $k$ ) [7].

This paper presents a new algorithm for maximum likelihood estimation of the covariance matrix (and therefore the delay-Doppler characteristics) of fast-fading mobile communication channels. Importantly, our algorithm explicitly makes use of the structural constraints. Key features of the algorithm include joint estimation of the channel mean and covariance, and applicability to a general class of wide-sense stationary (WSS) channels.

## 2. CHANNEL MODEL

### Channel Response

Consider a discrete equivalent baseband model in which the complex-valued time-varying channel, or reflectance process,  $f_{k,\epsilon}$ , represents the effect at time  $k$ , for reflections with a path delay  $\epsilon$ . Ignoring the average delay in the analysis, the observed signal is

$$z_k = \sum_{\epsilon=0}^{L-1} x_{k-\epsilon} f_{k,\epsilon} + w_k \quad (1)$$

where  $L$  is the length of the finite impulse response (FIR) channel, or the extent of the radar target,  $x_k$  is the known transmitted signal, and  $w_k$  is the additive noise introduced at the receiver.

Writing the observations for  $k = 0, \dots, N-1$  in vector notation,

$$\mathbf{z} = \mathbf{X}\mathbf{f} + \mathbf{w} \quad (2)$$

where the matrix of channel inputs is

$$\mathbf{X} = \begin{bmatrix} x_0 & \cdots & 0 & & x_{-L+1} & \cdots & 0 \\ \vdots & \ddots & \vdots & \cdots & \vdots & \ddots & \vdots \\ 0 & \cdots & x_{N-1} & & 0 & \cdots & x_{N-L} \end{bmatrix}$$

and  $\mathbf{F} = [f_{0,0}, \dots, f_{N-1,0}, \dots, f_{N-1,L-1}]^T$ <sup>1</sup>. The time-varying channel (or reflectance) process,  $\mathbf{F}$  is seen to be two-dimensional in that its elements are characterized both by the time index  $k$ , and the delay index  $\epsilon$ .

When a line-of-sight path or specular (stable) reflections exist between the transmitter and receiver, the channel is no longer zero-mean. Thus  $\mathbf{F} = \bar{\mathbf{F}} + \tilde{\mathbf{F}}$ , where  $\tilde{\mathbf{F}}$  is the zero-mean time-varying component and  $\bar{\mathbf{F}}$  is the mean component, constant over the observation interval,  $N$ . Here  $\bar{\mathbf{F}}$  is a  $NL \times 1$ , but contains only  $L$  independent parameters. For convenience, we also define the  $L \times 1$  vector  $\bar{\mathbf{G}} = (\mathbf{I} \otimes \mathbf{e}^T) \bar{\mathbf{F}}$ , and the corresponding  $N \times L$  matrix of channel inputs  $\mathbf{Y} = \mathbf{X}(\mathbf{I} \otimes \mathbf{1})$ , where  $\mathbf{e}^T = [1, 0, \dots, 0]$  is an  $1 \times N$  unit vector,  $\mathbf{1} = [1, \dots, 1]^T$  is an  $N \times 1$  vector of ones,  $\otimes$  is the Kronecker product operator, and  $\mathbf{I}$  is the  $L \times L$  identity matrix.

### Channel Covariance

The dimensionality of the channel impulse response is reflected in the structure of the covariance matrix

$$\begin{aligned} \mathbf{R} &= E[\tilde{\mathbf{F}}\tilde{\mathbf{F}}^H] \\ &= \begin{bmatrix} \mathbf{R}_{0,0} & \cdots & \mathbf{R}_{0,L-1} \\ \vdots & & \vdots \\ \mathbf{R}_{L-1,0} & \cdots & \mathbf{R}_{L-1,L-1} \end{bmatrix} \end{aligned} \quad (3)$$

which consists of  $L \times L$  blocks of  $N \times N$  matrices,  $\mathbf{R}_{\epsilon_1, \epsilon_2}$  which represent the covariance between taps (or reflectors) at delays  $\epsilon_1$  and  $\epsilon_2$ . For the radar target, where scatterers are assumed to behave independently (i.e. uncorrelated scatterers (US)) [5], the off-diagonal matrices will all be zero. However, for the communication channel model, the inclusion of the transmitter and receiver pulse shapes in the equivalent channel response,  $f_{k,\epsilon}$ , means that this is not the case.

When the statistics of the fading or reflectance process are wide-sense stationary (WSS) (in the dimension indexed by  $k$ ), the covariance matrices  $\mathbf{R}_{\epsilon_1, \epsilon_2}$  are Toeplitz. The overall matrix is then Hermitian symmetric and block-Toeplitz. The set of Hermitian block-Toeplitz matrices is denoted here by  $\mathbb{T}_{N,L}$ .

The Hermitian block-Toeplitz channel covariance ma-

<sup>1</sup>The transpose operator is denoted  $(\cdot)^T$ , and  $(\cdot)^H$  denotes a Hermitian transpose.

trix,  $\mathbf{R} \in \mathbb{T}_{N,L}$ , may be written as

$$\mathbf{R} = \sum_{m=1}^M r_m \mathbf{Q}_m \quad (4)$$

where  $r_m$  are the values of the real and imaginary components of elements of  $\mathbf{R}$ . There are  $M = 2NL^2 - L^2$  independent parameters,  $r_m$ . The channel covariance matrix is (by definition) positive semidefinite. This manifests itself as a highly nonlinear constraint on the parameters,  $r_m$ .

Assuming additive white Gaussian noise (AWGN) at the receiver, the channel covariance,  $\mathbf{R}$  is related to the  $N \times N$  observation covariance matrix,  $\mathbf{R}_z = E[\tilde{\mathbf{z}}\tilde{\mathbf{z}}^H]$  by

$$\mathbf{R}_z = \mathbf{X}\mathbf{R}\mathbf{X}^H + \sigma_w^2 \mathbf{I} \quad (5)$$

where  $\sigma_w^2$  is the variance of the observation noise, and the observation is  $\mathbf{z} = \bar{\mathbf{z}} + \tilde{\mathbf{z}}$ , where  $\bar{\mathbf{z}} = \mathbf{X}\bar{\mathbf{F}} = \mathbf{Y}\bar{\mathbf{G}}$  is the mean response.

### 3. MAXIMUM LIKELIHOOD CHANNEL ESTIMATION

To adequately identify the channel, we require estimates for the vector of channel tap means,  $\bar{\mathbf{G}}$ , and the matrix of channel tap covariances,  $\mathbf{R}$ . It is important that the estimates maximize the likelihood over the set of admissible structured matrices  $\mathbf{R} \in \mathbb{T}_{N,L}$ .

It is easily shown that maximizing the likelihood function for the channel model of Section 2 is the same as maximizing the following expression

$$\Phi(\bar{\mathbf{G}}, \mathbf{R}) = -\ln \det \mathbf{R}_z - \text{tr} \{ \mathbf{R}_z^{-1} \mathbf{S} \} \quad (6)$$

where the *sample covariance matrix*,  $\mathbf{S} = (\mathbf{z} - \mathbf{Y}\bar{\mathbf{G}})(\mathbf{z} - \mathbf{Y}\bar{\mathbf{G}})^H$  is a function of the mean channel  $\bar{\mathbf{G}}$ , and  $\mathbf{R}_z$  is a function of the channel covariance  $\mathbf{R}$  as given above in (5). Here,  $\text{tr} \{ \cdot \}$  denotes the trace operator.

Note that the likelihood, and hence  $\Phi(\bar{\mathbf{G}}, \mathbf{R})$ , is only defined when  $\mathbf{R}_z$  is strictly positive definite.

**Lemma 1** When  $\bar{\mathbf{G}}$  is given by

$$\bar{\mathbf{G}} = (\mathbf{Y}^H \mathbf{R}_z^{-1} \mathbf{Y})^{-1} \mathbf{Y}^H \mathbf{R}_z^{-1} \mathbf{z} \quad (7)$$

the first differential of the likelihood objective (6) is

$$d\Phi = \text{tr} \{ \mathbf{X}^H \mathbf{R}_z^{-1} (\mathbf{S} - \mathbf{R}_z) \mathbf{R}_z^{-1} \mathbf{X} d\mathbf{R} \} \quad (8)$$

**Proof** The first differential of the objective function (6) is [8]

$$d\Phi = -d(\ln \det \mathbf{R}_z) - \text{tr} \{ d(\mathbf{R}_z^{-1}) \mathbf{S} \}$$

Now, the first term is given by

$$d(\ln \det \mathbf{R}_z) = \text{tr} \{ \mathbf{R}_z^{-1} d\mathbf{R}_z \}$$

and the differential of an inverse is given by [8, pg 151]

$$d(\mathbf{R}_z^{-1}) = -\mathbf{R}_z^{-1} d\mathbf{R}_z \mathbf{R}_z^{-1}$$

Thus

$$\begin{aligned} d\Phi &= -\text{tr} \{ \mathbf{R}_z^{-1} d\mathbf{R}_z \} + \text{tr} \{ \mathbf{R}_z^{-1} d\mathbf{R}_z \mathbf{R}_z^{-1} \mathbf{S} \} \\ &\quad + 2 \text{tr} \{ \mathbf{R}_z^{-1} \mathbf{Y} d\bar{\mathbf{G}} (\mathbf{z} - \mathbf{Y}\bar{\mathbf{G}})^H \} \\ &= \text{tr} \{ \mathbf{R}_z^{-1} (\mathbf{S} - \mathbf{R}_z) \mathbf{R}_z^{-1} d\mathbf{R}_z \} \\ &\quad + 2 \text{tr} \{ (\mathbf{z} - \mathbf{Y}\bar{\mathbf{G}})^H \mathbf{R}_z^{-1} \mathbf{Y} d\bar{\mathbf{G}} \} \end{aligned} \quad (9)$$

Substituting (7) into (9) gives (8). ■

Unfortunately, due to non-linearities in (8) and the need for a positive definite solution, it is infeasible to obtain an analytic maximum likelihood solution for the covariance matrix using (8) by setting  $d\Phi = 0$ . We now present our main result in the following theorem which leads us to our recursive algorithm in Section 4 for finding an admissible maximum likelihood solution.

**Theorem 1** *The sequence of covariance matrices,  $\{\mathbf{R}_i\}$ , and channel tap means,  $\{\bar{\mathbf{G}}_i\}$ , generated by the following iterative equations (10) – (12), monotonically increases in likelihood*

$$\mathbf{R}_i = \mathbf{R}_{i-1} + \alpha_i (\tilde{\mathbf{R}}_i - \mathbf{R}_{i-1}) \quad (10)$$

$$\mathbf{R}_{z,i} = \mathbf{X} \mathbf{R}_i \mathbf{X}^H + \sigma_w^2 \mathbf{I} \quad (11)$$

$$\bar{\mathbf{G}}_i = (\mathbf{Y}^H (\mathbf{R}_{z,i})^{-1} \mathbf{Y})^{-1} \mathbf{Y}^H (\mathbf{R}_{z,i})^{-1} \mathbf{z} \quad (12)$$

where  $\alpha_i > 0$  is an arbitrarily small stepsize, and where

$\tilde{\mathbf{R}}_i = \sum_{n=1}^M \tilde{r}_{n,i} \mathbf{Q}_n$ , for  $\tilde{r}_{n,i}$  satisfying the following set of equations, for  $m = 1, \dots, M$

$$\begin{aligned} &\sum_{n=1}^M \text{tr} \{ \mathbf{X}^H \mathbf{R}_{z,i-1}^{-1} \mathbf{X} \mathbf{Q}_n \mathbf{X}^H \mathbf{R}_{z,i-1}^{-1} \mathbf{X} \mathbf{Q}_m \} \tilde{r}_{n,i} \\ &= \text{tr} \{ \mathbf{X}^H \mathbf{R}_{z,i-1}^{-1} (\mathbf{S}_{i-1} - \sigma_w^2 \mathbf{I}) \mathbf{R}_{z,i-1}^{-1} \mathbf{X} \mathbf{Q}_m \} \end{aligned} \quad (13)$$

where  $\mathbf{R}_{z,i-1} = \mathbf{X} \mathbf{R}_{i-1} \mathbf{X}^H + \sigma_w^2 \mathbf{I}$  and  $\mathbf{S}_{i-1} = (\mathbf{z} - \mathbf{Y}\bar{\mathbf{G}}_{i-1})(\mathbf{z} - \mathbf{Y}\bar{\mathbf{G}}_{i-1})^H$ . The initial  $\mathbf{R}_0$  must be positive definite Hermitian block-Toeplitz (e.g.  $\mathbf{R}_0 = \mathbf{I}$ ).

**Proof** From (8), consider the differential of the likelihood objective function at iteration  $i$

$$\begin{aligned} d\Phi_i &= \text{tr} \{ \mathbf{X}^H \mathbf{R}_{z,i-1}^{-1} (\mathbf{S}_{i-1} - \mathbf{R}_{z,i-1}) \mathbf{R}_{z,i-1}^{-1} \mathbf{X} d\mathbf{R}_i \} \\ &= \text{tr} \{ \mathbf{X}^H \mathbf{R}_{z,i-1}^{-1} (\mathbf{S}_{i-1} - \mathbf{R}_{z,i-1} \\ &\quad - (1/\alpha_i) \mathbf{X} d\mathbf{R}_i \mathbf{X}^H) \mathbf{R}_{z,i-1}^{-1} \mathbf{X} d\mathbf{R}_i \} \\ &\quad + (1/\alpha_i) \text{tr} \{ \mathbf{X}^H \mathbf{R}_{z,i-1}^{-1} \mathbf{X} d\mathbf{R}_i \mathbf{X}^H \mathbf{R}_{z,i-1}^{-1} \mathbf{X} d\mathbf{R}_i \} \end{aligned} \quad (14)$$

In order to prove that the recursion (10) increases the likelihood (i.e. when  $d\mathbf{R}_i = \alpha_i (\tilde{\mathbf{R}}_i - \mathbf{R}_{i-1})$ ), we now proceed to show that the second term in (14) is positive, and the first term is zero.

The second term in (14) may be written

$$\begin{aligned} &(1/\alpha_i) \text{tr} \{ \mathbf{X}^H \mathbf{R}_{z,i-1}^{-1} \mathbf{X} d\mathbf{R}_i \mathbf{X}^H \mathbf{R}_{z,i-1}^{-1} \mathbf{X} d\mathbf{R}_i \} \\ &= (1/\alpha_i) \text{tr} \{ \mathbf{X}^H \mathbf{A} \mathbf{A}^H \mathbf{X} d\mathbf{R}_i \mathbf{X}^H \mathbf{A} \mathbf{A}^H \mathbf{X} d\mathbf{R}_i \}; \\ &\quad \mathbf{R}_{z,i-1}^{-1} = \mathbf{A} \mathbf{A}^H \text{ since p.d.} \\ &= (1/\alpha_i) \text{tr} \{ \mathbf{A}^H \mathbf{X} d\mathbf{R}_i \mathbf{X}^H \mathbf{A} \mathbf{A}^H \mathbf{X} d\mathbf{R}_i \mathbf{X}^H \mathbf{A} \} \\ &= (1/\alpha_i) \text{tr} \{ \mathbf{B} \mathbf{B}^H \}; \\ &\quad \mathbf{B} = \mathbf{A}^H \mathbf{X} d\mathbf{R}_i \mathbf{X}^H \mathbf{A} \text{ and } d\mathbf{R}_i = d\mathbf{R}_i^H \\ &> 0 \end{aligned}$$

Now, before considering the first term in (14), consider (13), which can be written

$$\begin{aligned} &\text{tr} \{ \mathbf{X}^H \mathbf{R}_{z,i-1}^{-1} (\mathbf{S}_{i-1} - \sigma_w^2 \mathbf{I}) \mathbf{R}_{z,i-1}^{-1} \mathbf{X} \mathbf{Q}_m \} \\ &\quad - \text{tr} \{ \mathbf{X}^H \mathbf{R}_{z,i-1}^{-1} \mathbf{X} \tilde{\mathbf{R}}_i \mathbf{X}^H \mathbf{R}_{z,i-1}^{-1} \mathbf{X} \mathbf{Q}_m \} = 0 \\ &\text{tr} \{ \mathbf{X}^H \mathbf{R}_{z,i-1}^{-1} (\mathbf{S}_{i-1} - \sigma_w^2 \mathbf{I} - \mathbf{X} \mathbf{R}_{i-1} \mathbf{X}^H \\ &\quad - (1/\alpha_i) \mathbf{X} d\mathbf{R}_i \mathbf{X}^H) \mathbf{R}_{z,i-1}^{-1} \mathbf{X} \mathbf{Q}_m \} = 0; \\ &\quad \text{since } \tilde{\mathbf{R}}_i = \mathbf{R}_{i-1} + (1/\alpha_i) d\mathbf{R}_i \\ &\text{tr} \{ \mathbf{X}^H \mathbf{R}_{z,i-1}^{-1} (\mathbf{S}_{i-1} - \mathbf{R}_{z,i-1} \\ &\quad - (1/\alpha_i) \mathbf{X} d\mathbf{R}_i \mathbf{X}^H) \mathbf{R}_{z,i-1}^{-1} \mathbf{X} \mathbf{Q}_m \} = 0 \\ &\sum_{m=1}^M \text{tr} \{ \mathbf{X}^H \mathbf{R}_{z,i-1}^{-1} (\mathbf{S}_{i-1} - \mathbf{R}_{z,i-1} \\ &\quad - (1/\alpha_i) \mathbf{X} d\mathbf{R}_i \mathbf{X}^H) \mathbf{R}_{z,i-1}^{-1} \mathbf{X} \mathbf{Q}_m \} d\mathbf{r}_m = 0 \\ &\text{tr} \left\{ \mathbf{X}^H \mathbf{R}_{z,i-1}^{-1} (\mathbf{S}_{i-1} - \mathbf{R}_{z,i-1} \right. \\ &\quad \left. - (1/\alpha_i) \mathbf{X} d\mathbf{R}_i \mathbf{X}^H) \mathbf{R}_{z,i-1}^{-1} \mathbf{X} \sum_{m=1}^M \mathbf{Q}_m d\mathbf{r}_m \right\} = 0 \end{aligned}$$

Since  $d\mathbf{R}_i = \sum_{m=1}^M \mathbf{Q}_m d\mathbf{r}_m$ , the first term in (14) is zero. ■

**Remark 1** *Theorem 1 utilizes the inverse iteration argument of [1]. However, this new result is applicable when the process of interest is not necessarily zero-mean and is observed via convolution in additive noise. We have included estimation of both the real and imaginary parts of each of these channel parameters which arise from the baseband model. Since the result is not restricted to zero-mean and uncorrelated scatterer models, it leads to a more generally applicable algorithm than the circulant extension algorithm of [5].*

#### 4. RECURSIVE ALGORITHM

Theorem 1 provides us with a recursive algorithm for maximum likelihood channel mean and structured covariance estimates. However, in order to perform an iteration (10)–(12), we must first solve (13). At each iteration  $i$ , this can be done by forming a vector  $\mathbf{x} = [r_{1,i}, \dots, r_{M,i}]^T$  and an  $M \times 1$  vector  $\mathbf{b}$  with elements given by the RHS of (13) for  $m = 1, \dots, M$ . Now the set of equations in (13) for  $m = 1, \dots, M$  can be written in the form  $\mathbf{A}\mathbf{x} = \mathbf{b}$ , where we can now solve for  $\mathbf{x}$ . It is easily shown that at each iteration  $\mathbf{A}$  is positive definite, and therefore efficient algorithms can be employed in the solution.

This new recursive estimation algorithm is in fact a linearized gradient algorithm, as can be seen by the linear equation (13). The formulation can easily be extended for multiple observations.

**Remark 2** For the directly observable case presented in [1], it was sufficient to confine the estimates of the structured covariance matrix at each iteration to the positive definite region (by appropriate choice of the stepsize) to obtain an admissible maximum likelihood solution. Note that for the case presented here however, that  $\mathbf{R}_z$  (5) may be positive definite even when the estimate of  $\mathbf{R}$  is not. Since the maximum of the objective (6) may occur in this region, gradient algorithms may not be guaranteed to find an admissible ( $\mathbf{R} \in \mathbf{T}_{N,L}$ ) maximum of the objective function.

##### Example A

Our new recursive algorithm was first tested on a zero-mean US channel. The channel was simulated with  $L = 2$  independent equal power fading taps with a Jakes' Doppler spectrum. The signal to noise ratio (SNR) was nominally chosen to be 10 dB. The dimension of the covariance matrix was  $NL = 50$  and 75 samples of the channel output were used in the estimation (representing a multiple observation factor of 3). The stepsize at each iteration,  $\alpha_i$ , was chosen to confine the corresponding estimate  $\mathbf{R}_i$  to the positive definite region.

Figure 1 shows the progression of the objective maximization with respect to computational effort. Also shown is the progression of the algorithm of [5], using a factor of 2 for the circulant extension. The scaling of the curves relative to the computational effort was based on counts of floating point operations in MATLAB for unoptimized code in both cases, and therefore the figure is only indicative of a performance comparison.

Importantly, Figure 1 shows that the restriction of the

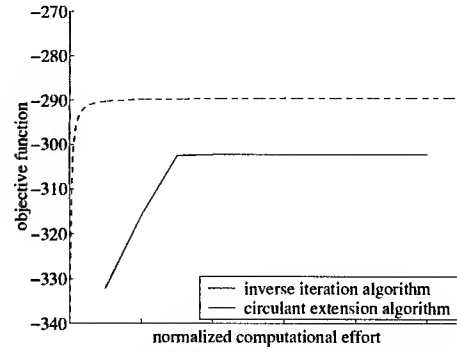


Figure 1: Maximization of the likelihood objective relative to computational effort, estimates  $\mathbf{R}_i$  restricted to the positive definite region

estimates  $\mathbf{R}_i$  to the positive definite region at each iteration may result in trapping the algorithm at the positive definite boundary when a solution with greater likelihood exists in the admissible region.

**Remark 3** The algorithm in [5] for US channels exploits the circulant extension property of Toeplitz matrices [9], and has been shown to be an instance of the expectation-maximization (EM) algorithm. With sensible initialization, this algorithm maintains a positive definite estimate of  $\mathbf{R}$ . However, due to the augmentation of the Toeplitz matrix to a circulant matrix, the estimation problem is modified, and conditions for convergence to an admissible maximum of (6) have not yet been established.

##### Modification of the gradient

To pursue the admissible maximum likelihood solution, modification of the gradient is required to allow movement tangential to the positive definite boundary whilst maintaining the positive definite constraint on the estimates  $\mathbf{R}_i$ . Due to the complexity of the relationship between the positive definite constraint and the parameters  $r_m$ , no obvious modification strategy is apparent.

A simple modification we have found is to replace the set of linear equations for calculating  $\tilde{\mathbf{R}}_i$  (13) with

$$\begin{aligned} & \sum_{n=1}^M \text{tr} \{ \mathbf{R}_{i-1}^{-1} \mathbf{Q}_n \mathbf{R}_{i-1}^{-1} \mathbf{Q}_m \} \tilde{r}_{n,i} \\ & = \text{tr} \{ \mathbf{X}^H \mathbf{R}_{z,i-1}^{-1} \mathbf{S}_{i-1} \mathbf{R}_{z,i-1}^{-1} \mathbf{X} \mathbf{Q}_m \} \end{aligned} \quad (15)$$

It is an unproven conjecture of this paper that this modified algorithm converges to an admissible maximum likelihood solution for the structured covariance matrix.

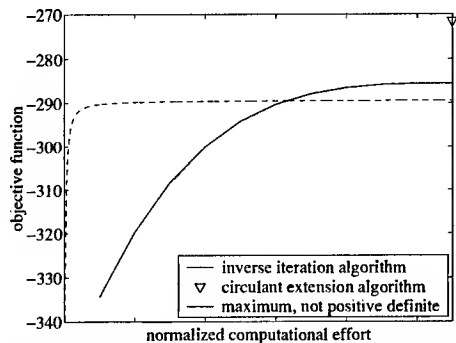


Figure 2: Maximization of the likelihood objective relative to computational effort, modified gradient

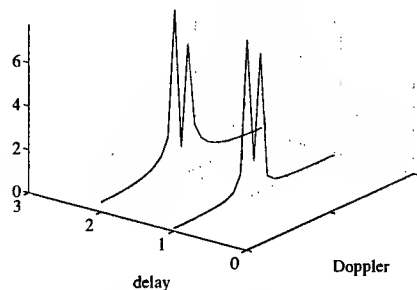


Figure 3: Delay-Doppler profile of the simulated channel

### Example B

The experiment of Example A was repeated using the modified gradient described above. Note the smooth trajectory of the modified algorithm, suggesting that the algorithm is no longer trapped prematurely. Also shown is the likelihood ( $\nabla$ ) obtained for the structured covariance matrix estimate without the positive definite constraint.

Figure 3 shows the delay-Doppler profile of the simulated channel. Figures 4 and 5 show the corresponding estimates of the delay Doppler spectrum. Improvement can be obtained using more data (with correspondingly more computational effort) and/or higher SNR. Further trials show that the modified gradient algorithm is robust in estimating the channel mean, with good mean estimates and negligible impact on the covariance estimate and delay-Doppler profile.

### REFERENCES

- [1] J. P. Burg, D. G. Luenberger, and D. L. Wenger, "Estimation of structured covariance matrices," in *Proceedings of the IEEE*, vol. 70, pp. 963-974, Sept. 1982.
- [2] M. I. Miller and D. L. Snyder, "The role of likelihood and entropy in incomplete-data problems: Applications to estimating point-process intensities and Toeplitz con-

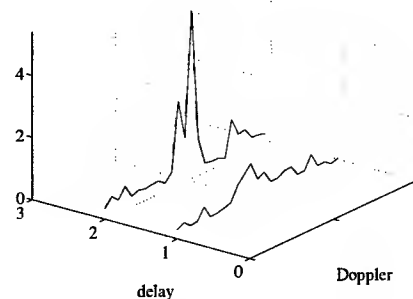


Figure 4: Estimated delay-Doppler profile, circulant extension algorithm

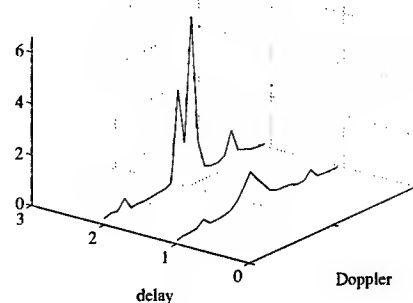


Figure 5: Estimated delay-Doppler profile, modified gradient algorithm

- strained covariances," *Proceedings of the IEEE*, vol. 75, pp. 892-907, July 1988.
- [3] A. Dembo, C. L. Mallows, and L. A. Shepp, "Embedding nonnegative definite Toeplitz matrices in nonnegative definite circulant matrices, with application to covariance estimation," *IEEE Trans. on Information Theory*, vol. 35, pp. 1206-1212, Nov. 1989.
- [4] L. M. Davis, R. J. Evans, and E. Polak, "Maximum likelihood estimation of positive definite Hermitian Toeplitz matrices using Outer Approximations," in *Proc. of IEEE Workshop on Statistical Signal and Array Processing (SSAP'98)*, (Portland, OR, USA), pp. 49-52, Sept. 1998.
- [5] D. L. Snyder, J. A. O'Sullivan, and M. I. Miller, "The use of maximum likelihood estimation for forming images of diffuse radar targets from delay-Doppler data," *IEEE Trans. on Information Theory*, vol. 35, pp. 536-548, Nov. 1989.
- [6] L. M. Davis, I. B. Collings, and R. J. Evans, "Estimation of LEO satellite channels," in *Int. Conf. on Information, Communications and Signal Processing (ICICS'97)*, vol. 1, (Singapore), pp. 15-19, Sept. 1997.
- [7] H. L. Van Trees, *Detection Estimation and Modulation Theory*, vol. III. Wiley, 1971.
- [8] J. R. Magnus and H. Neudecker, *Matrix Differential Calculus with Applications in Statistics and Econometrics*. Wiley, 1988.
- [9] R. M. Gray, "Toeplitz and circulant matrices: II," tech. rep., Center for Systems Research, Stanford University, Apr. 1977.

# ENHANCED SPACE-TIME CAPTURE PROCESSING FOR RANDOM ACCESS CHANNELS

*Alexandr M. Kuzminskiy, Kostas Samaras, Carlo Luschi and Paul Strauch*

Bell Laboratories, Lucent Technologies  
Unit 1, Pagoda Park, Westmead Drive  
Swindon, Wiltshire SN5 7YT, UK  
ak9@lucent.com

## ABSTRACT

The problem of maximizing the throughput in a Random Access Channel (RACH) in a TDMA-based system is addressed. A general analysis of a Slotted ALOHA system is presented which shows that a possibility to recover more than one user in a RACH collision can significantly improve system performance. Three capture algorithms based on semi-blind space-time filtering are proposed. Their efficiency compared to the conventional (power) training-based capture algorithm, is demonstrated by means of simulations in a GSM(EDGE) system. The best results are obtained for a multistage version of the training-like algorithm based on the Least Squares (LS) estimation of space-time filter coefficients.

## 1. INTRODUCTION

Cellular mobile communication systems such as the GSM make use of RACH in order to enable the initial access of the mobile stations to the network. Packet radio networks (like GPRS and EGPRS) also make use of similar channels called Packet Random Access Channels (PRACH) not only for the initial access but also during the call since channels are allocated to users on a demand basis, rather than permanently (as in circuit switched GSM). The random access mechanism used in these systems is based on the Slotted ALOHA principle [1]. The throughput in a slotted ALOHA random access channel in a TDMA-system system can be improved by using capture effects. Most capture models in TDMA-based systems rely on power capture [2] and not more than one of colliding packets can be recovered. Specifically, when more than one packet arrive at the receiver simultaneously only one of them can be captured at the receiver given that its power exceeds a specified threshold. Capture of more than one packet in a collision of many, leads to performance enhancement. We start from the general analysis of a Slotted ALOHA system with capture. We show that the throughput can be increased significantly if a nonzero probability of capture of more than one packet in

a collision is assumed. Then we propose three capture algorithms based on semi-blind space-time filtering. The first one is based on a multistage procedure where each stage exploits the conventional LS estimator with ability to capture at most one of the colliding packets. The second algorithm is based on a training-like (TL) approach [5,6] that allows us to introduce a nonzero probability to recover more than one user in a collision of many using an one stage procedure. The third one is a combination of the multistage and training-like algorithm. Simulations in a GSM(EDGE) context are presented, which demonstrate the superior performance of the multiple capture algorithms compared to the LS estimator.

## 2. CAPTURE EFFECTS IN A SLOTTED ALOHA SYSTEM

In order to demonstrate the performance enhancement due to space-time capture processing we consider a simple S-ALOHA system with a finite population of users,  $N$ . A generalization of the model described in [2,3] is adopted where the input load of the system is described by the probability of packet arrival denoted as  $p_0$ . Each of the users (terminals) generates single packet messages with probability  $p_0$ . A discrete time system is considered and transmissions of packets occur only at the boundaries between two time slots. If the transmission of a packet is not successful the terminal is backlogged and makes an attempt to retransmit the packet in the next time slot with retransmission probability  $p_r$ . The capture ability of the channel is described by the capture matrix  $C = [c(i, j)]$ , where  $c(i, j)$  denotes the probability that there are  $i$  successfully received packets given that there are  $j$  packet transmission attempts in the same time slot ( $0 \leq i, j \leq N$ ). It is assumed that all transmitting terminals are aware of the outcome of their transmissions before the end of the time slot through an ideal feedback (downlink) channel. The state of the system can be described by the number  $n$  of backlogged terminals ( $0 \leq n \leq N$ ).

The steady state behavior of this discrete time Markovian system is determined by the  $(N + 1) \times (N + 1)$  transition



probabilities matrix  $\Pi = [\pi_{n,m}]$ , where  $\pi_{n,m}$  is the probability that the state of the system (population of backlogged terminals) is  $m$  during time slot  $t+1$ , given that during time  $t$  the state was  $n$ . The adopted model allows us to express these transition probabilities as follows:

$$\pi_{n,m} = \sum_{i=0}^{N-n} \sum_{j=0}^n \sum_{k=\max\{n-m+i-j, 0\}}^{\min\{n-m+i, i\}} \binom{N-n}{i} p_0^i (1-p_0)^{N-n-i} \binom{n}{j} p_r^j (1-p_r)^{n-j} c(k, i) c(n-m+i-k, j). \quad (1)$$

The expression for the transition probabilities in [3] is a special case of (1) when the capture matrix of the system becomes:

$$c(i, j) = \begin{cases} 1 - q_j, & i = 0 \\ q_j, & i = 1 \\ 0, & i > 1 \end{cases}, \quad (2)$$

where  $q_j$  is the probability that one out of  $j$  transmitted packets is successfully received. A semi-analytical approach has been followed for the calculation of the transition probabilities. The elements of the capture probability matrix  $C$ , for the purposes of this paper, have been calculated through simulation. In particular the elements  $c(i, j)$  with  $1 \leq i \leq M_1$ ,  $1 \leq j \leq M_2$  (typical values  $M_1 = 3$ ,  $M_2 = 5$ ) are calculated via simulation, and  $c(0, j) = 1 - \sum_{i=1}^{M_1} c(i, j)$  for  $1 \leq j \leq M_2$ . Furthermore,  $c(0, 0) = 1$  and  $c(i, j) = 0$  for all other  $(i, j)$ .

The steady state distribution  $\mathbf{P} = \{P_k\}_{k=0}^N$  of the number of backlogged users is given as the solution to the following problem [4]:

$$\mathbf{P} \cdot \Pi = \mathbf{P} \quad (3)$$

under the constraint:

$$\sum_{k=0}^N P_k = 1. \quad (4)$$

As a performance metric the average number of successfully transmitted packets per time slot has been chosen, which is referred to as the average throughput  $\bar{S}$ . The average throughput can be calculated as follows:

$$\bar{S} = \sum_{(n,m)} S(n) P_n, \quad (5)$$

where  $S(n)$  denotes the number of successful packet transmissions when the system is in state  $n$  and can be calculated by:

$$S(n) = \sum_{m=0}^N \sum_{i=0}^{N-n} \sum_{j=0}^n (n-m+i) \cdot \pi_{i,j}. \quad (6)$$

A possibility to improve the system performance by means of capture effects is illustrated in Figure 1, where the average system throughput as a function of the retransmission probability is plotted for no capture and the ideal capture ( $\bar{S} = Np_0$ ) where  $N = 10$  and  $p_0 = 0.2$ . One can see the significant gap between these two boundary cases, which can be filled by curves corresponding to algorithms with multiple capture ability.

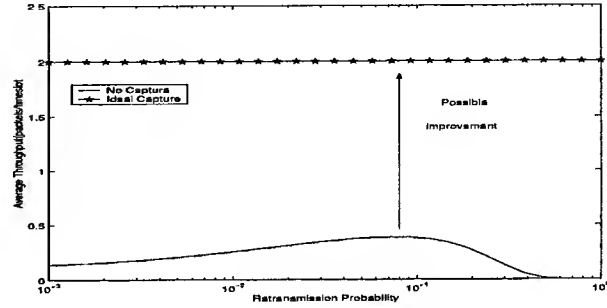


Figure 1: Slotted ALOHA throughput performance for the boundary cases

### 3. MULTIPLE CAPTURE PROBLEM FORMULATION

The model of a RACH collision is shown in Figure 2. The main assumptions are:

- 1) all colliding signals and Co-Channel Interference (CCI) have the known structure of a timeslot (GSM, for example) and they are received synchronously,
- 2) all signals are from the same finite alphabet (FA)  $\{a_h, h = 1, \dots, J\}$  and all of them have the same training sequence which is different compared to the CCI training sequence,
- 3) channel coding is used (successful capture can be detected by means of parity check),
- 4) multiple antenna is used at the receiver (space-time interference rejection filtering can be applied),
- 5) propagation channels for all colliding signals are stationary over the whole time slot (coefficients of a space-time filter can be adjusted by means of off-line algorithms).

The main difficulty to recover more than one user is that the training data for all access packets in one cell is the same. This means that training-based algorithms cannot be directly applied for multiple capture reception. Blind techniques could be applicable, but short burst nature of Slotted ALOHA systems makes it unrealistic because of the finite amount of data effects [7]. A possibility to address this problem by means of semi-blind space-time filtering algorithms is studied in this paper.

**Note:** The important feature of the considered problem is that some probability of access failure can be acceptable

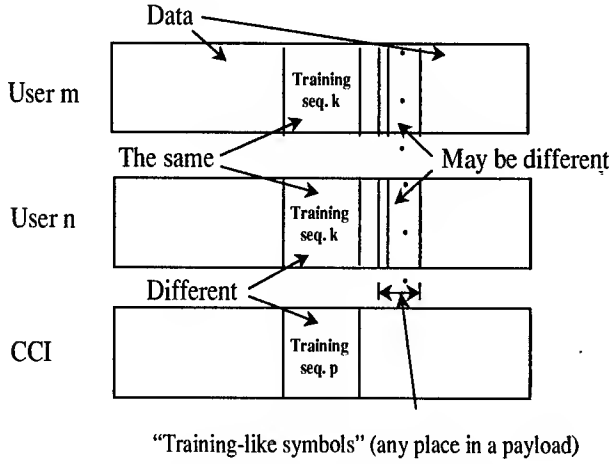


Figure 2: Model of a RACH collision

for RACH systems. Thus, solutions without proven ability to recover all colliding signals in every time slot may be useful.

#### 4. MULTIPLE CAPTURE ALGORITHMS

##### 4.1. Multistage algorithm

A multistage processing based on cancellation of the recovered signals from the received signal at successive stages has been considered for different applications, for example in [8,9]. A possible way to implement this technique in the considered problem is presented in Figure 3 (two stages are shown for simplicity). The conventional LS algorithm is used at each stage in the Space-time Filter. The possible number of stages can be found from the applicability condition (misadjustment) for the Noise Canceller [10]:

$$\frac{(\text{Number of stages} - 1) * \text{Length of channels}}{\text{Number of information symbols in a timeslot}} < 1.$$

The advantage of this algorithm is that more than one user may be captured if the first stage is successful. The disadvantage is that no signals can be recovered if there is no capture at the first stage. We refer to this straightforward algorithm as the MLS (multistage LS) and consider it as a reference point for the enhanced algorithms introduced in the next two subsections.

##### 4.2. One stage training-like algorithm

According to a general TL approach [5], our proposal is to use a few information symbols in the payload as an extension of the training sequence. These symbols may be different for different users. Thus, the enlarged training sequences may be linearly independent and the LS estimator based on these TL sequences can be applied. In Figure 2

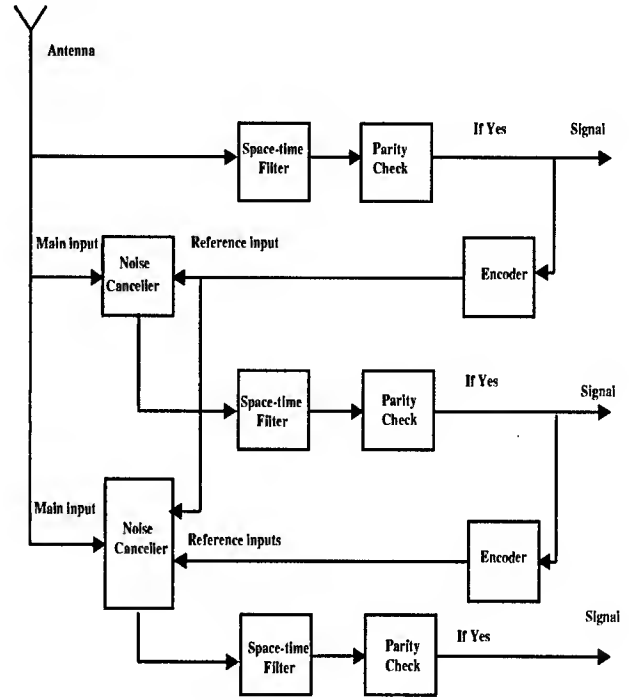


Figure 3: Structure of the MLS algorithm

these information symbols are indicated as the TL symbols. The coefficients of the space-time filters and signal estimations corresponded to the TL sequences can be found for the FA signals using the following training-like LS (TLIS) algorithm:

- form the  $J^{N_{TL}}$  TL sequences

$$\{s\}_m^{TL} = \{s(n_1)s(n_2)...s(n_{N_T})\} \quad (7)$$

where  $s(n_i)$ ,  $i = 1, ..., N_T$  are the training symbols,  $\{\tilde{s}_m(m_1)\tilde{s}_m(m_2)...\tilde{s}_m(m_{N_{TL}})\}$  are all  $J^{N_{TL}}$  possible sequences of the FA signal of the length  $N_{TL}$ ;  $n_i$  and  $m_j$  are the positions of the known and TL symbols ( $n_i$ ,  $i = 1...N_T$  are known,  $m_j$ ,  $j = 1...N_{TL}$  must be selected);

- calculate the LS estimations of the weight vectors using the TL sequences

$$\mathbf{W}_m = (\hat{\mathbf{R}} + \delta \mathbf{I})^{-1} \hat{\mathbf{P}}_m, \quad m = 1...J^{N_{TL}}, \quad (8)$$

where

$$\hat{\mathbf{R}} = \sum_{i=1}^{N_T} \mathbf{X}(n_i) \mathbf{X}^*(n_i) + \sum_{j=1}^{N_{TL}} \mathbf{X}(m_j) \mathbf{X}^*(m_j); \quad (9)$$

$$\hat{\mathbf{P}}_m = \sum_{i=1}^{N_T} s^*(n_i) \mathbf{X}(n_i) + \sum_{j=1}^{N_{TL}} \tilde{s}_m^*(m_j) \mathbf{X}(m_j); \quad (10)$$

where  $\mathbf{X}$  is the vector of input signals,  $\delta$  is the regularization coefficient [11] for the conventional LS estimator

which usually is chosen to be close to the variance of the noise;

- select  $M_1$  weight vectors which minimize the distance from the FA  $Q_m$

$$\hat{\mathbf{W}}_j = \mathbf{W}_{m_j}, \quad m_j = \arg \min_{m=1 \dots J N_{TL}} Q_m, \quad j = 1, \dots, M_1, \quad (11)$$

$$Q_m = \sum_{n=1}^{N_s} \min_h (|a_h - \mathbf{W}_m^* \mathbf{X}(n)|), \quad (12)$$

where  $N_s$  is the number of symbols in a time slot;

- calculate signal candidates

$$s_j(n) = \hat{\mathbf{W}}_j^* \mathbf{X}(n), \quad (13)$$

- apply parity check to each signal candidate and accept different signal candidates with the positive parity check as the captured packets.

The drawback of this solution is that the number of the TL sequences grows exponentially with the number of the TL symbols. Thus, only a small number of the TL symbols can be implemented. Certainly, in this situation we cannot guarantee the possibility to recover all signals in a collision in each timeslot. Nevertheless, according to the Note in Section 3 this is not necessary in the considered problem. We have introduced a multiple capture ability in an one stage procedure and, in Section 5, we will demonstrate the performance improvement for only two TL symbols in the GSM(EDGE) environment.

#### 4.3. Multistage training-like algorithm

Capture ability can be additionally improved by means of multistage processing similar to that presented in Section 4.1 when the TLLS algorithm is used instead of the LS estimator. We refer to this algorithm as the MTLLS (multistage TLLS).

### 5. SIMULATION RESULTS

Two antennas receiving in a typical GSM ( $J = 2$ ) urban scenario TU50 is assumed, where SNR=35dB and SIR=6dB. In all cases a space-time filter with five coefficients in each channel is used. For each time slot, the transmitted bits are obtained by channel encoding of one data block. The channel coding scheme includes a (34,28) systematic cyclic redundancy check (CRC) code (which accepts 28 bits at the input and provides 6 parity check bits at the output), and a (3,1,5) convolutional code (rate 1/3, constraint length 5).

A possibility to capture more than one user in a collision for the TLLS algorithm is illustrated in Figure 4, where the typical curves for the selection criteria (distance from the

FA) are shown for  $N_{TL} = 4$  (16 TL sequences for the binary FA) in the case of two colliding users ( $M_2 = 2$ ). All situations are presented in Figure 4: no capture, one of two users is captured, and two of two users are captured. Our goal is to estimate probabilities of these events for different  $M_2$  and then to calculate the system performance according to the semi-analytical procedure presented in Section 2. The capture simulation results (estimated probabilities  $p_i$ ,  $i = 1, 2, 3$  to recover one, two or three colliding packets) are given in Table 1 for the conventional LS algorithm (at most one signal can be captured), for the TLLS with  $N_{TL} = 2$ , and for the MTLLS with the same  $N_{TL}$ .

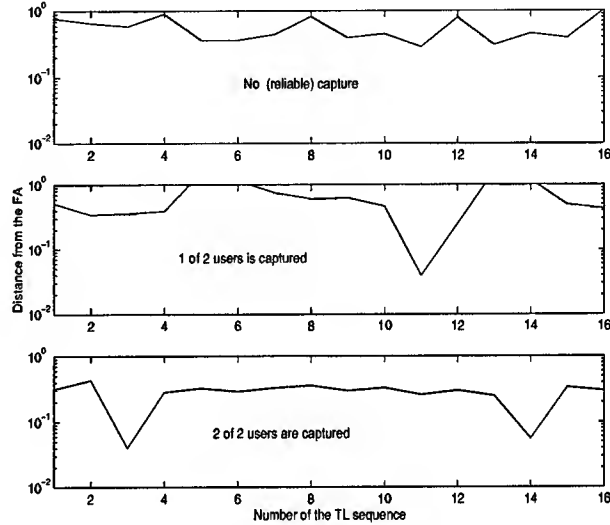


Figure 4: Illustration of the selection step in the TLLS for  $N_{TL} = 4$

Table 1. Estimated probabilities to capture one/two/three packets in a collision of one/.../five packets

$M_2$	$p_i$	Algorithm			
		LS	TLLS	MLS	MTLLS
1	$p_1$	1	1	1	1
	$p_2$	0	0	0	0
	$p_3$	0	0	0	0
2	$p_1$	0.87	0.47	0.01	0
	$p_2$	0	0.51	0.86	0.96
	$p_3$	0	0	0	0
3	$p_1$	0.68	0.56	0.21	0.20
	$p_2$	0	0.30	0.14	0.09
	$p_3$	0	0.03	0.033	0.59
4	$p_1$	0.54	0.56	0.32	0.32
	$p_2$	0	0.17	0.13	0.23
	$p_3$	0	0.01	0.1	0.19
5	$p_1$	0.41	0.47	0.30	0.35
	$p_2$	0	0.1	0.09	0.15
	$p_3$	0	0	0.02	0.05

The corresponding curves for the average system throughput as a function of the retransmission probability are shown in Figure 5 for the conditions indicated in Section 2. One can see the significant performance improvement for the enhanced algorithms, especially for the MTLLS, compared to the conventional LS estimator even for only two TL symbols.

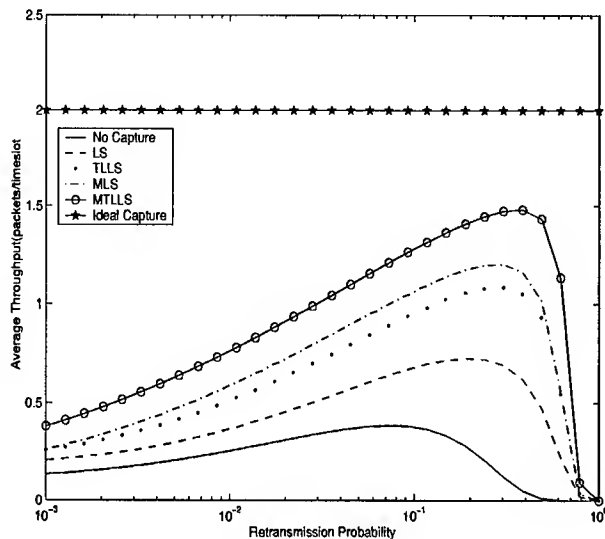


Figure 5: Slotted ALOHA throughput performance for different algorithms

## 6. CONCLUSION

It has been shown analytically that a possibility to recover more than one user in a RACH collision can significantly improve system performance. A semi-analytical approach has been proposed to evaluate the average throughput over a Slotted ALOHA system with multiple capture. Three semi-blind space-time filtering algorithms with multiple capture ability have been presented. Their efficiency compared to the training-based algorithm with a power capture has been demonstrated in a GSM(EDGE) environment.

## 7. REFERENCES

- [1] L. G. Roberts, "ALOHA packet system, with and without slots and capture", *ACM Computer Communication Review*, vol. 5, no. 2, pp. 28-42, Apr. 1975.
- [2] C. Namislo, "Analysis of Mobile Radio Slotted ALOHA Networks", *IEEE Journal on Selected Areas in Communications*, vol. SAC-2, no. 4, pp. 583-588, Jul. 1984.
- [3] J.J. Metzner, "Comments on a widely used capture model for Slotted ALOHA", *IEEE Transactions on Communications*, vol. 44, no. 4, p. 419, Apr. 1996.

- [4] W. Feller, "An introduction to probability theory and its applications", Wiley, 1968.
- [5] A.M.Kuzminskiy, D.Hatzinakos, "Semi-blind estimation of spatio-temporal filter coefficients based on a training-like approach", *IEEE Signal Processing Letters*, vol. 5, n. 9, pp. 231-233, Sept. 1998.
- [6] A.M.Kuzminskiy, P.Strauch, "Space-time filtering with suppression of asynchronous co-channel interference", to be published in *Proc. AS-SPCC*, 2000.
- [7] A.M.Kuzminskiy, "Finite amount of data effects in spatio-temporal filtering for equalization and interference rejection in short burst wireless communications", to be published in *Signal Processing*, vol. 80, n. 10, 2000.
- [8] G.J.M.Janssen, "BER and outage performance of a dual signal receiver for narrowband BPSK modulated co-channel signals in a Rician fading channel", in *Proc. PIMRC*, pp. 601-606, 1994.
- [9] A.M.Kuzminskiy, D.Hatzinakos, "Multistage semi-blind spatio-temporal processing for short burst multiuser SDMA systems", in *Proc. 32nd Asilomar Conf. on Signals, Systems and Computers*, pp. 1887-1891, 1998.
- [10] B.Widrow, J.M.McCool, M.G.Larimore, C.R.Johnson, Jr., "Stationary and nonstationary learning characteristics of the LMS adaptive filters", *Proc. IEEE*, vol. 64, pp. 1151-1162, Aug. 1976.
- [11] Y.I.Abramovich, "Controlled method for adaptive optimization of filters using the criterion of maximum SNR", *Radio Engineering and Electronic Physics*, vol.26, n.3, pp.87-95, 1981.

# ASYMMETRIC SIGNALING CONSTELLATIONS FOR PHASE ESTIMATION

Trasapong Thaiupathump, Charles D. Murphy and Saleem A. Kassam

Department of Electrical Engineering  
University of Pennsylvania  
Philadelphia, PA 19104  
e-mail: kassam@ee.upenn.edu

## ABSTRACT

In digital communication systems, most commonly used signaling constellations are symmetric. Without a pilot tone or known training sequence, an arbitrary phase rotation cannot be identified from a symmetric constellation. The standard approach to overcome the phase ambiguity is to use differential encoding. In this paper, we introduce the notion of using an asymmetric constellation instead of a symmetric constellation with differential encoding. The absolute phase of an asymmetric constellation can be determined using blind statistics of processed channel outputs. Through simulation and analysis, we study the trade-offs between asymmetry and other features of a constellation, such as, data rate, power, and symbol separation.

## 1. INTRODUCTION

A symmetric constellation has the property that blind processing is unable to identify an arbitrary rotation of symbols. Synchronization with the phase of the transmitted carrier may be done by using pilot tones or known training sequences. In blind system, without a pilot tone or training sequence, the receiver must rely on statistics of channel outputs to recover the phase of the received signal. All of the commonly-used symbol constellations - PAM, PSK, QAM, and others - are symmetric when the symbols are equiprobable. Blind statistics of these constellations cannot produce an absolute phase estimate. To overcome the phase ambiguity, a mapping between the data and the symbols has to be invariant to an unknown reference phase. A simple method is to use differential encoding. Since each symbol is used to determine two symbol transitions, a symbol decision error will usually result in two transition errors. The penalty incurred by differential encoding is well characterized as a 2-3 dB loss in SNR [2], [3].

In this paper, we introduce the notion of using an asymmetric constellation. The absolute phase of an asymmetric constellation can be estimated using blind statistics of processed channel outputs. We discuss symmetry, asymmetry, and how to design asymmetric constellations and absolute phase estimators. Through simulation and analysis, we study the performance of various absolute phase estimators and the trade-offs between asymmetry and other features of a constellation.

## 2. SYMMETRY BREAKING

$M$ -ary PAM, QAM, and PSK are the most often encountered symmetric constellations. A symmetric constellation may be rendered asymmetric by changing the symbol values and/or the symbol probabilities.

Consider an  $M$ -ary constellation with  $M = 2^m$  equiprobable i.i.d. symbols. The data rate (in bits/symbol) or entropy of the constellation is

$$H(S) = - \sum_{i=0}^{M-1} p_i \log_2 p_i = m \quad (1)$$

where  $p_i$  is the probability of symbol  $i$ . If the number of symbols and the symbol locations are to remain unchanged, an asymmetric constellation can be obtained by adjusting the symbol probabilities. Because the symbols in the asymmetric constellation are no longer equiprobable, the data rate of the constellation is strictly lower than that of the corresponding symmetric constellation. This is a trade-off of data-rate for asymmetry.

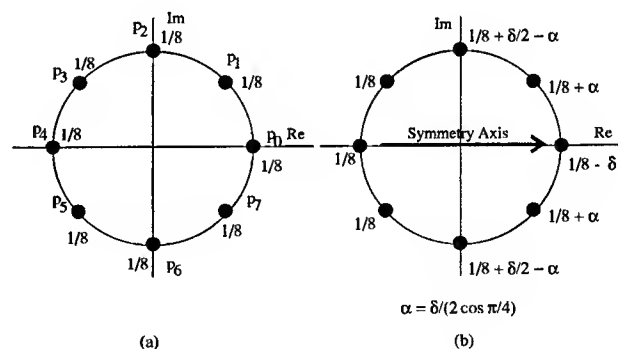


Figure 1: Symmetric and Asymmetric 8-PSK (Manipulation of the Symbol Probabilities)

Fig.1(a) illustrates an 8-PSK constellation with equiprobable symbols  $\sqrt{A} \cdot e^{j\pi i/4}$ ,  $i = 0, \dots, 7$ , constant transmitted power  $A$ , and a data rate of 3 bps. On the right is an asymmetric 8-PSK obtained by manipulating the symbol probabilities. The value of  $p_0$  has been reduced by a small  $\delta$ ,  $0 < \delta \leq 1/8$ . To maintain  $\sum_{i=0}^7 p_i = 1$  and a zero DC value, the probabilities of some other symbols have also been changed. Since the symbols in the second constellation

are no longer equiprobable, the data rate of the constellation is strictly less than 3 bps. Figure 2 shows the exact reduction in entropy as a function of  $\delta$  for  $0 \leq \delta \leq 0.12$ .

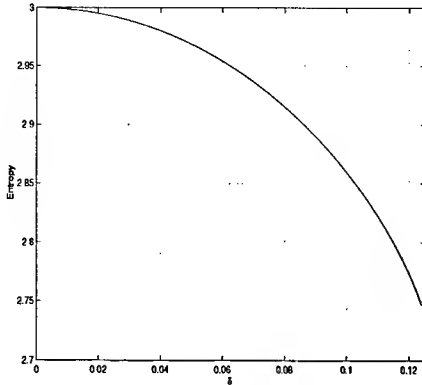


Figure 2:  $\delta$  vs.  $H(S)$  for the Asymmetric 8-PSK

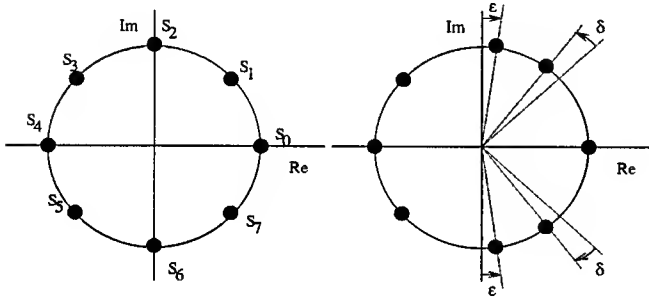


Figure 3: Symmetric and Asymmetric 8-PSK (Symbol Relocation)

In Fig.3, another alternative to introducing asymmetry is by relocating some of the original 8-PSK constellation points without changing the equal probability assigned to 8 points. With symbol probability and power unchanged, the symbol  $s_1$  is rotated counterclockwise by  $\delta$  radians. In order to maintain the zero-mean condition, symbols  $s_2$ ,  $s_6$ , and  $s_7$  must also be relocated. The symbols  $s_2$ ,  $s_6$ , and  $s_7$  are rotated by  $-\epsilon$ ,  $+\epsilon$ , and  $-\delta$  radians, respectively, where  $\epsilon = \sin^{-1}\{\cos(\pi/4) \cdot (1 - \cos(\delta) + \sin(\delta))\}$ . However, introducing asymmetry by moving some symbols closer together will cause more erroneous symbol decisions. With perfect phase estimation, the union bound of the error rate is

$$P_e \leq \frac{1}{2} \left[ Q \left( \sqrt{\frac{2E_s}{N_0}} \sin \left( \frac{\pi}{8} + \frac{\delta}{2} \right) \right) + Q \left( \sqrt{\frac{2E_s}{N_0}} \sin \left( \frac{\pi}{8} - \frac{(\delta + \epsilon)}{2} \right) \right) + Q \left( \sqrt{\frac{2E_s}{N_0}} \sin \left( \frac{\pi}{8} + \frac{\epsilon}{2} \right) \right) \right]$$

$$+ Q \left( \sqrt{\frac{2E_s}{N_0}} \sin \left( \frac{\pi}{8} \right) \right) \right]. \quad (2)$$

For small  $\delta$ , the error rate performance can be very close to that of coherent symmetric 8-PSK constellation.

### 3. ABSOLUTE PHASE ESTIMATION

#### 3.1. Maximum Likelihood Approach

Consider the transmission of nonequiprobable  $M$ -PSK signals over an AWGN channel. The  $M$ -PSK symbol has the complex form  $s_i = \sqrt{A}e^{j2\pi i/M}$ ,  $i = 0, 1, \dots, M-1$ , where  $\sqrt{A}$  denotes the constant signal power. The transmitted symbol  $x[n] = s_i$  with probability  $p_i$ . The corresponding received sequence is then

$$r[n] = x[n]e^{j\phi} + w[n] \quad n = 0, 1, \dots, N-1 \quad (3)$$

where  $w[n]$  is a sample of zero-mean complex white Gaussian noise and  $\phi \in (0, 2\pi)$  is an arbitrary phase introduced by the channel. For the assumed AWGN model, the pdf of  $r[n]$  can be modeled as a mixture of  $M$  distributions

$$p(r[n]; \phi) = \sum_{i=0}^{M-1} (p_i) \cdot f_i(r[n]; \phi) \quad (4)$$

where

$$f_i(r[n]; \phi) = \frac{1}{2\pi\sigma^2} \exp \left( -\frac{|r[n] - s_i e^{j\phi}|^2}{2\sigma^2} \right). \quad (5)$$

Then the pdf of the sequence  $\mathbf{r}$  is

$$p(\mathbf{r}; \phi) = \prod_{n=0}^{N-1} \left[ \sum_{i=0}^{M-1} (p_i) \cdot f_i(r[n]; \phi) \right]. \quad (6)$$

The MLE of  $\phi$  is the value that maximizes the likelihood function in Eq.(6). In general, the derivative of  $\ln p(\mathbf{r}; \phi)$  with respect to  $\phi$  does not reduce to a simple form. The MLE of  $\phi$  can be obtained numerically by using iterative maximization procedures. The difficulty with the use of these numerical methods is that in general the point found may not be the global maximum but possibly only a local maximum or even a local minimum.

A simpler alternative likelihood method of finding the absolute phase is based on the use of the phase statistics. We may express the unknown phase as  $\phi = \phi_0 + k \cdot (\frac{2\pi}{M})$  where  $k = \text{integer}$ ,  $0 \leq k \leq M-1$  and  $0 \leq \phi_0 \leq 2\pi/M$ . Let  $\theta[n]$  be the phase angle of the received sequence  $r[n]$ . The  $\phi_0$  is obtained first by

$$\hat{\phi}_0 = \frac{1}{M} \bar{\psi} \quad (7)$$

where

$$\bar{\psi} = \text{angle} \left( \frac{1}{N} \sum_{n=0}^{N-1} e^{jM\theta[n]} \right) \quad (8)$$

is the mean phase angle of the received sequence after each phase angle has been multiplied by  $M$ . Then, the maximum likelihood method can be applied to find the correct value

of integer  $k$ . Using the estimate  $\hat{\phi}_0$ , the complex plane is divided into slices  $q_i$ ,  $i = 0, 1, \dots, M-1$  bounded by phase angles  $\{\pi/M + \hat{\phi}_0 + i \cdot (2\pi/M), i = 0, 1, \dots, M-1\}$ . Although the nonequiprobable symbols do cause the optimum symbol-by-symbol decision boundaries to change at the receiver, these angular decision boundaries are close to being optimum for small  $\delta$ . Let  $l_i$  be the number of points in the received data sequence that fall in each region  $q_i$ . Then, we are able to obtain the integer  $k$  that maximizes the likelihood function defined as

$$\hat{k} = \arg \max_k p(\mathbf{n}; k) \quad (9)$$

where  $p(\mathbf{n}; k)$  can be modeled as a multinomial distribution with  $\mathbf{n} = [n_0 \ n_1 \ \dots \ n_{M-1}]$  and  $n_i = l_{(i+k) \pmod M}$

$$\begin{aligned} p(\mathbf{n}; k) &= \frac{N!}{n_0! n_1! \dots n_{M-1}!} p_0^{n_0} p_1^{n_1} \dots p_{M-1}^{n_{M-1}} \\ &= N! \prod_{i=0}^{M-1} \frac{p_i^{n_i}}{n_i!}. \end{aligned} \quad (10)$$

This is equivalent to finding the integer  $k$  that maximizes the log-likelihood function

$$\ln p(\mathbf{n}; k) = \ln \prod_{i=0}^{M-1} p_i^{n_i} = \sum_{i=0}^{M-1} n_i \ln p_i. \quad (11)$$

Therefore, by using simple bin statistics, the absolute phase estimate is

$$\hat{\phi} = \hat{\phi}_0 + \hat{k} \cdot \left( \frac{2\pi}{M} \right). \quad (12)$$

To make the correct decision in estimating  $k$ , the estimates  $\hat{p}_i$  should be close to their true value  $p_i$ . Finding the sample size  $N$  to generate reliable estimates of the  $p_i$  requires the joint probabilities that the  $\hat{p}_i$  lie within some  $\epsilon$  intervals centered on the correct values. Using Chebyshev's Inequality, we can roughly determine the number of required samples  $N$  such that the estimate  $\hat{p}_i$  is within  $\epsilon$  of its correct value  $p_i$  with probability  $1 - \tau$ .

$$P\{|\hat{p}_i - p_i| \leq \epsilon\} \geq 1 - \frac{\sigma_i^2}{\epsilon^2} = 1 - \tau \quad (13)$$

where  $\hat{p}_i = n_i/N$  with variance  $\sigma_i^2 = \{p_i(1 - p_i)\}/N$ . Setting  $\epsilon = \delta/P$ , the required  $N$  is given by

$$N = \frac{p_i(1 - p_i)P^2}{\tau\delta^2}. \quad (14)$$

For the asymmetric setting shown in Fig.1(b), the most likely incorrect  $\hat{k}$  are the correct value of  $k$  offset by  $\pm 2 \pmod 8$ . The two largest probability symbols  $p_1$  and  $p_7$  in Fig.1 appear to be the most critical values to consider. From Eq.(14), setting  $p_i = p_1 = 1/8 + \alpha = 1/8 + \delta/\sqrt{2}$ , we obtain

$$N = \frac{(7 + 8\sqrt{2}\delta - 32\delta^2)P^2}{64\tau\delta^2}. \quad (15)$$

In Fig.4, we plot  $(N\tau/P^2)$  as a function of  $\delta$ . As an example, setting  $\tau = 0.1$  and  $P = 2\sqrt{2}$  which corresponds to setting  $\epsilon$  to half of the difference between the largest and the

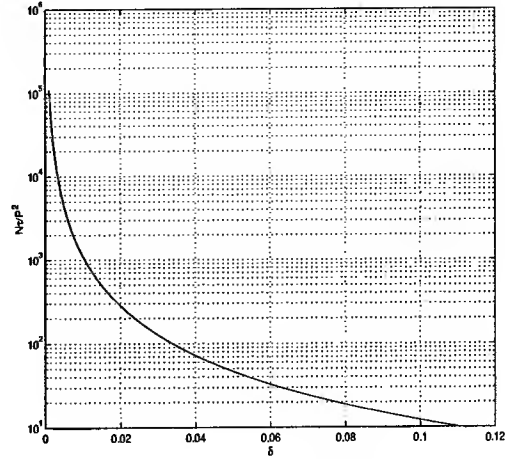


Figure 4:  $(N\tau/P^2)$  as a function of  $\delta$

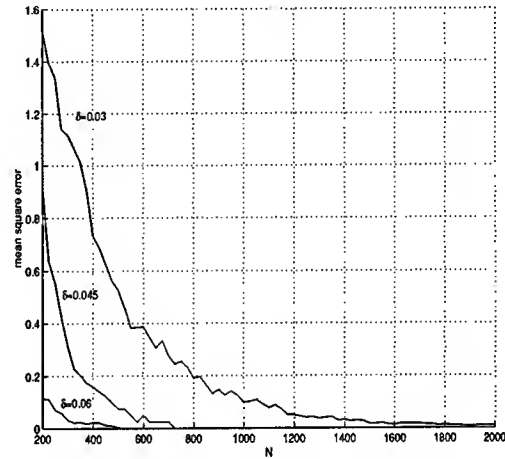


Figure 5: Comparison of MSE of phase estimation with different  $\delta$ .

second largest values of symbol probabilities, at  $\delta = 0.06$ ,  $(N\tau/P^2) \approx 32$  which gives  $N \approx 2,600$  samples. Figure 5 shows simulation results for the MSE of phase estimation as a function of  $N$  for  $\sigma^2 = 0.01$  and  $A = 1$ . At the same level of MSE performance, the sample size needed for  $\delta = 0.03$  is approximately 4 times larger than the sample size needed for  $\delta = 0.06$ . While the MSE performance includes contributions from both estimation of  $\phi_0$  and estimation of  $k$ , the relative dependence of  $N$  on  $\delta$  (i.e. the factor by which  $N$  increases for decreasing  $\delta$ ) is captured well by the approximation of Fig.4.

Figure 6 illustrates the error probability performance of the various approaches. The bottom dashed line shows the error probability performance of the coherent symmetric 8-PSK. The top curve shows the error probability of symmetric 8-DPSK. The stars show the simulated error rates of asymmetric 8-PSK with  $\delta = 0.06$  ( $H(S) = 2.9542$ ) and  $\sqrt{A} = 1$ . The symbols are rotated by an unknown constant phase  $\phi \in (0, 2\pi)$  radians and further distorted by AWGN. Statistics of 1,000 samples are used to estimate the

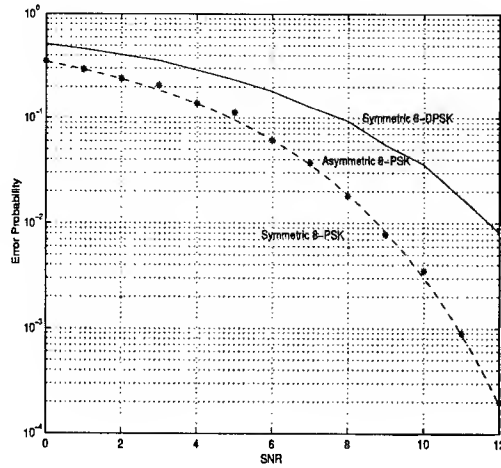


Figure 6: Asymmetric 8-PSK, Symmetric 8-PSK, and Symmetric 8-DPSK Error Rate Comparison

absolute phase angle. As shown in Fig.6, the performance of asymmetric 8-PSK is close to 3 dB better than that for symmetric 8-DPSK at large SNR. Note that the data rate of asymmetric constellation is less than that of symmetric 8-DPSK by approximately 1.52%. Thus, in order to make a meaningful comparison of these two modulation methods, we should allow the 8-PSK symmetric constellation to use some form of encoding with rate 0.985. However, to obtain a coding gain of 3 dB, the rate will have to be significantly lower. Thus we conclude that when we have large enough sample size for phase estimate, the performance of asymmetric 8-PSK can be close to that for coherent symmetric 8-PSK constellation.

### 3.2. Nonparametric Methods

Without any prior knowledge on probability distribution and exact locations of symbol values, nonparametric or distribution-free methods can be used to estimate the absolute phase rotation.

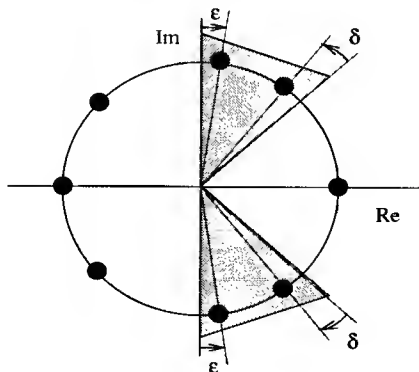


Figure 7: An absolute phase estimation scheme for Asymmetric 8-PSK obtained by changing symbol locations

For an asymmetric constellation obtained by changing

the symbol locations as in Fig.3, a simple and effective scheme for phase estimation is based on noting that at the correct zero angle, roughly half of the samples will fall in two angular regions bounded by  $\pi/4$  and  $\pi/2$  and by  $-\pi/4$  and  $-\pi/2$ , shown as two shaded regions in Fig.7. The absolute phase can be estimated by searching for the angle that gives the maximum number of points in these two angular bins. This scheme works well in the presence of some noise, however, at high SNR, this scheme is only able to obtain the estimate within  $\epsilon$  of the correct phase angle. We can further search for the angle within this range that gives the minimum mean square error from the center angle between these search sectors.

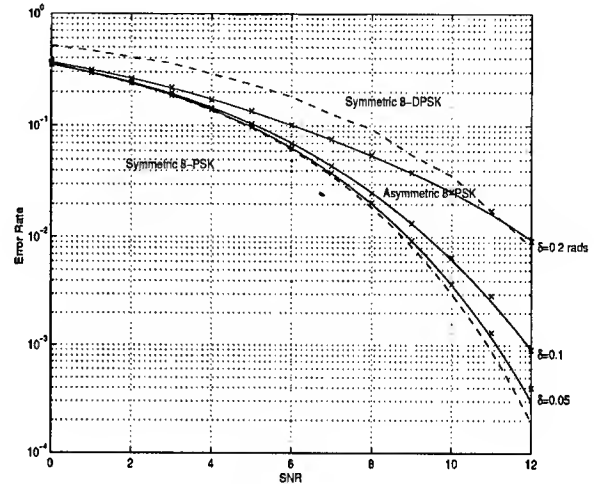


Figure 8: Comparison of the noise performances of asymmetric 8-PSK constellations obtained by changing symbol locations, with different  $\delta$  (symbol relocation case).

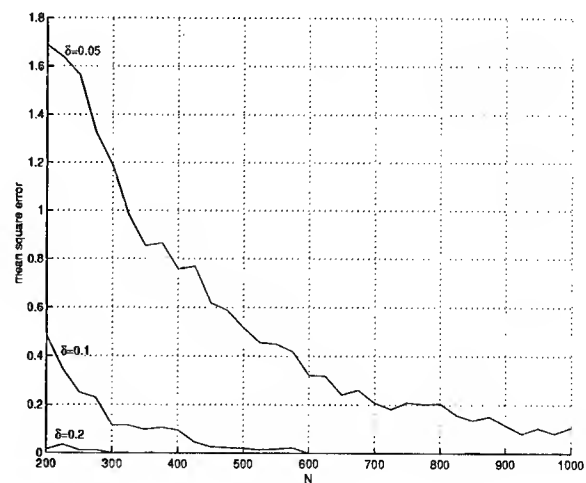


Figure 9: Comparison of MSE of phase estimation with different  $\delta$  (symbol relocation).

Figure 8 shows the error probability performance for symmetric 8-DPSK and asymmetric 8-PSK obtained by



changing symbol locations, with different values of  $\delta$ . The top and bottom dashed lines show the error rate performance of symmetric 8-DPSK and symmetric 8-PSK, respectively. The solid lines show the union bound of the error performance and x marks show the simulated results. Results are based on 1,000 equiprobable i.i.d. symbols rotated by an unknown phase  $\phi$  and further corrupted by AWGN.

Figure 9 illustrates the MSE of the phase estimate with different values of  $\delta$ , assuming that the equiprobable i.i.d. symbols are rotated by an unknown phase  $\phi$  and further distorted by AWGN with variance 0.01. Fig.8 illustrates that with large  $\delta$ , the estimate converges to the absolute phase faster than with low  $\delta$ . However, with larger  $\delta$ , some symbol points are relocated closer to adjacent symbol points which will cause more erroneous symbol decisions.

The shape of the mask that is used in estimating the absolute phase is not unique. The mask shown in Fig.7 is just an example. We can use different masks bounded by some different angular boundaries, such as, half-plane shape bounded by  $-\pi/2$  and  $\pi/2$  angles. The properties of a good mask shape are straightforward. It should give a maximum number of points at the correct angle and the number of points in the mask should fall when it rotates away from the correct angle. Sensitivity analysis can be used to evaluate the performance of the mask.

#### 4. DISCUSSION

A symmetric constellation may be rendered asymmetric by changing the symbol values and/or the symbol probabilities. Between these two methods of introducing asymmetry to existing symmetric 8-PSK, manipulating symbol probabilities will certainly cause some reduction of the number of data bit transmitted per symbol and some additional complexity in encoding/decoding process to obtain the asymmetric probability arrangement. For the second asymmetric arrangement, the symbol probabilities are remain unchanged, so the data rate is the same as that of a symmetric constellation, without additional complexity in a coding process.

#### 5. CONCLUSION

An asymmetric constellation is introduced as an alternative to regular symmetric constellation with differential encoding. Without the use of a pilot tone or known training sequence, the absolute phase of received symbols can be estimated blindly from asymmetric constellation using simple statistics of the received symbols. By introducing asymmetry to existing symmetric constellation, the absolute phase recovery function is obtained at the cost of very small reduction in entropy and/or minimum distance. Both the asymmetry of a constellation and the phase recovery function may be considered as choices much as symbol separation, the number of bits transmitted per symbol, or power, providing new tools for constellation design.

#### 6. REFERENCES

- [1] C.D. Murphy, *Blind Equalization of Linear and Non-linear Channels*, Ph.D. Thesis, University of Pennsyl-

vania, 1999.

- [2] R.D. Gitlin, J.F. Hayes, and S.B. Weinstein, *Data Communications Principles*, New York: Plenum Press 1992.
- [3] J.G. Proakis, *Digital Communications*, New York: McGraw-Hill 1995.
- [4] G.J. Foschini and R.D. Gitlin, "Optimization of Two-Dimensional Signal Constellations in the Presence of Gaussian Noise," *IEEE Trans. on Communications*, Vol. COM-22, No. 1, January 1974.
- [5] D.G. Forney et al., "Efficient Modulation for Band-Limited Channels," *IEEE J., Selected Areas in Communications*, Vol. SAC-2, pp. 632-647, August 1984.

# A CONVEX SEMI-BLIND COST FUNCTION FOR EQUALIZATION IN SHORT BURST COMMUNICATIONS

*Kelvin K. Au and Dimitrios Hatzinakos*

Department of Electrical and Computer Engineering,  
University of Toronto, Toronto, Ontario, Canada, M5S 3G4  
Tel: (416) 978-1613, Fax: (416) 978-4425  
{aukar,dimitris}@comm.toronto.edu

## ABSTRACT

In short burst wireless communications, a training sequence is incorporated in each burst for the receiver to adjust the equalizer coefficients. However, when the amount of training symbols is less than the spatial-temporal equalizer tap weights, conventional least-square technique may not provide good MSE performance. Blind methods, on the other hand, may not achieve equalization in a short burst. A regularized semi-blind algorithm was proposed previously by Kuzminskiy et al. to overcome this problem but local minima exist in the algorithm. A convex cost with training symbols as the equalizer constraint is proposed in this paper to avoid cost-dependent local minima. Furthermore, comparison with the regularized semi-blind algorithm suggests that the proposed algorithm achieves a lower MSE performance in the case of non-constant modulus signals such as 16-QAM signals.

## 1. INTRODUCTION

Conventional equalization techniques in wireless communications require transmission of training sequences. This represents a system overhead and effectively reduces the information rate. On the other hand, blind equalization algorithms do not require training. One of the most popular blind algorithms is the family of constant modulus algorithms (e.g. CMA 2-2 or Godard [2] algorithm, CMA 1-2 or Sato algorithm). There are several disadvantages in using the CMA family of algorithms. One of them is the existence of local minima. In situations where fractionally-spaced equalizer or antenna array are used, the Godard algorithm was shown to converge globally [3]. Unfortunately, this is not true for CMA 1-2 (Sato) algorithm which was demonstrated to have cost-dependent local minima in either case [4].

This work was supported by the Natural Sciences and Engineering Research Council of Canada (NSERC).

Another drawback of blind algorithms is the slow convergence and inability to achieve equalization in a short burst.

A regularized semi-blind algorithm was proposed in [1] which combined the LS and CM 1-2 costs. The ability to successfully equalize the channels with a spatial-temporal filter was demonstrated and thus offered the possibility of reducing the number of training symbols. However, local minima inherent to the cost exist. Using a convex cost will eliminate the possibility of convergence to cost-dependent local minima. Blind convex cost with equalizer tap-anchoring was introduced in [5, 6]. In this paper, we shall make use of the training sequence in conjunction with the blind convex cost [6] to formulate a new and more efficient semi-blind algorithm. Simulation results demonstrate the potential of the proposed algorithm for constant and non-constant modulus signals.

## 2. SPATIAL-TEMPORAL SIGNAL MODEL

We assume there are  $K$  users in the model. One of the user is the signal of interest. Without loss of generality, we shall denote the first user to be the desired signal. The remaining  $K-1$  signals are coming from nearby co-channel cells. At the base station receiver, an antenna array of  $M$  sensors is employed.

The data is processed in a burst of  $N$  symbols which are assumed to be received under a stationary environment. There are  $N_t$  training symbols in each burst and the starting position of the training sequence is  $N_s$ , which is assumed to be known. The transmitted signals undergo linear channels which are assumed to be FIR of length  $N_c$ . This assumption is valid when we have a finite delay spread. Equalization is necessary when the delay spread is larger than the symbol duration. The

received signal at the  $j$ -th sensor is given by:

$$y_j(n) = \sum_{i=1}^K c_{ij}^H x_i(n) + v_j(n) \quad (1)$$

for  $i = 1, \dots, K, j = 1, \dots, M$

$$\mathbf{c}_{ij} = [c_{ij}(0), \dots, c_{ij}(N_c - 1)]^T, \quad (2)$$

$$\mathbf{x}_i(n) = [x_i(n), \dots, x_i(n - N_c + 1)]^T, \quad (3)$$

where  $H$  denotes the conjugate transpose of a matrix and each  $c_{ij}(n)$  is a complex Gaussian random variable whose amplitude does not change over the duration of the burst. The noise  $v_j(n)$  is a complex circularly symmetric additive white Gaussian noise of variance  $\sigma_n^2$ .

Recall that the first user is the desired signal. The equalizer output for the signal of interest is given by:

$$z_1(n) = \mathbf{w}^H \mathbf{y}(n), \quad (4)$$

where  $\mathbf{y}(n) = [\mathbf{y}_1^T(n), \dots, \mathbf{y}_M^T(n)]^T$  and each  $\mathbf{y}_j(n) = [y_j(n), \dots, y_j(n - N_w + 1)]^T$ . The spatial-temporal equalizer taps are  $\mathbf{w} = [\mathbf{w}_1^T, \dots, \mathbf{w}_M^T]^T$  and each  $\mathbf{w}_j = [w_{j1}, \dots, w_{jN_w}]^T$ . The vector  $\mathbf{w}$  has a dimension of  $MN_w \times 1$ .

### 3. LS SOLUTION WITH FEW TRAINING SYMBOLS

When a burst of known symbols (training) is received, the method of least square can be used to obtain the spatial-temporal equalizer coefficients. The following equation is satisfied:

$$\hat{\mathbf{R}}\mathbf{w} = \hat{\mathbf{p}}, \quad (5)$$

where  $\hat{\mathbf{R}}$  is the time-averaged spatial-temporal auto-correlation matrix

$$\hat{\mathbf{R}} = \frac{1}{N_t} \sum_{n=N_s}^{N_s+N_t-1} \mathbf{y}(n)\mathbf{y}(n)^H, \quad (6)$$

and  $\hat{\mathbf{p}}$  is the time-averaged spatial-temporal cross-correlation matrix

$$\hat{\mathbf{p}} = \frac{1}{N_t} \sum_{n=N_s}^{N_s+N_t-1} x_1^*(n-d)\mathbf{y}(n) \quad (7)$$

for some delay  $d$ .

If the number of training symbols is fewer than the number of spatial-temporal equalizer coefficients  $N_w M$ ,  $\hat{\mathbf{R}}$  has null( $\hat{\mathbf{R}}$ ) =  $N_w M - N_t$ . Therefore there are many solutions to (5) which can be expressed as:

$$\mathbf{w} = \hat{\mathbf{R}}^+ \hat{\mathbf{p}} + \sum_{i=1}^{N_w M - N_t} v_i \mathbf{U}_i, \quad (8)$$

where  $\hat{\mathbf{R}}^+$  is the pseudo-inverse of  $\hat{\mathbf{R}}$ ,  $\mathbf{U}_i$ 's are a set of orthonormal basis of the null space of  $\hat{\mathbf{R}}$  and  $v_i$ 's are a set of coefficients. Equation (8) can be expressed compactly as:

$$\mathbf{w} = \hat{\mathbf{R}}^+ \hat{\mathbf{p}} + \mathbf{U}\mathbf{v}, \quad (9)$$

where  $\mathbf{U} = [\mathbf{U}_1, \mathbf{U}_2, \dots, \mathbf{U}_{N_w M - N_t}]$  and  $\mathbf{v} = [v_1, v_2, \dots, v_{N_w M - N_t}]^T$ .

The semi-blind algorithm in [1] tried to regularize the standard LS solution with the CM 1-2 cost to provide a better estimation of the equalizer coefficients in the case of  $N_t < N_w M$ . The algorithm minimizes the cost

$$J(\mathbf{w}) = \frac{1}{N_t} \sum_{n=N_s}^{N_s+N_t-1} |z_1(n) - x_1(n-d)|^2 + \rho \frac{1}{N} \sum_{n=1}^N (|z_1(n)| - R_1)^2 \quad (10)$$

where  $R_1 = E|a_n|^2 / E|a_n|$  with  $a_n$  being the alphabets in a signal constellation and  $0 < \rho < \infty$  is a regularized constant. We shall refer readers to [1] for details of the algorithm.

### 4. SEMI-BLIND EQUALIZATION BASED ON A CONVEX COST FUNCTION

#### 4.1. Background

Since cost-dependent local minima exist in the regularized semi-blind algorithm, there are two ways to avoid convergence to such minima: 1) devising a good initialization strategy of equalizer tap weights or 2) choosing alternative cost functions that are convex. In this paper, we are primarily interested in adopting a convex cost function in the problem of semi-blind equalization.

In [5] (and references therein), a convex cost function based on the  $l_\infty$  norm of an equalizer output was proposed in the context of blind equalization. The idea comes from the fact that the opening of the eye of the signal constellation is characterized by the intersymbol interference (ISI). Suppose the combined channel-equalizer response is  $c * w = h$ , the eye is opened when the magnitude of  $h(\delta)$  for some delay  $\delta$  dominates the rest of the coefficients,  $\sum_{i \neq \delta} |h(i)|$ . This is closely related to the  $l_1$  norm of the combined channel-equalizer response. In practice, however, we can never know the channel response explicitly. An equivalent but more useful formulation is using the  $l_\infty$  norm of the equalizer output [5, 6, 7]. In [6], the following cost is proposed:

$$J(\mathbf{w}) = \|\text{Re}(z(n))\|_\infty + \|\text{Im}(z(n))\|_\infty \quad (11)$$

with the constraint

$$\text{Re}(w_{jk}) + \text{Im}(w_{jk}) = 1. \quad (12)$$

Two remarks about (11) and (12) are in order:

1. The cost (11) is appropriate for square-type constellations such as 4-QAM, 16-QAM etc.
2. The constraint (12) anchors one of the equalizer taps. This is needed to avoid the all-zero equalizer coefficients which is a valid but trivial minimum to this type of convex cost function.

#### 4.2. Convex cost with training constraint

In this section, we propose a linear constraint to be used for the convex cost (11). We call it semi-blind because the linear constraint makes use of the small amount of known training symbols present in the received burst of data. The idea was essentially discussed in the previous section. When the number of training symbols is fewer than the spatial-temporal equalizer coefficients, the solution of the LS problem can be expressed as (and restated here):

$$\hat{\mathbf{R}}\mathbf{w} = \hat{\mathbf{p}} \quad (13)$$

$$\mathbf{w} = \hat{\mathbf{R}}^+ \hat{\mathbf{p}} + \mathbf{U}\mathbf{v}. \quad (14)$$

Equation (13) can be viewed as a constraint on the equalizer and can be adopted to replace the tap-anchoring technique. Hence (11) and (13) describe our semi-blind convex cost.

There are several properties of this semi-blind algorithm:

1. The semi-blind constraint (13) is linear. It can be thought of as a generalization of the tap-anchoring technique.
2. Because of the linear constraint, convexity of the cost (11) is still preserved.
3. Convexity of the cost (11) is established in a doubly infinite equalizer (ideal) setting and also in a finitely parameterized equalizer (practical) setting [6]. Therefore, using an FIR equalizer maintains convexity unlike the Godard cost function.
4. As in the case of the blind convex cost function, this kind of equalization technique leaves an unknown gain at the equalizer output [7]. Hence an automatic gain control (AGC) is needed to scale the output. This can be done with the knowledge of the known signal constellation.

#### 4.3. Implementation

Since  $l_\infty$  norm cannot be implemented in practice, we approximate the  $l_\infty$  norm with  $l_p$  norm for some large  $p$ :

$$\begin{aligned} J(\mathbf{w}) &= \|\text{Re}(z(n))\|_\infty + \|\text{Im}(z(n))\|_\infty \\ &\simeq \lim_{p \rightarrow \infty} \|\text{Re}(z(n))\|_p + \|\text{Im}(z(n))\|_p \\ &\simeq (E|\text{Re}(z(n))|^p)^{\frac{1}{p}} + (E|\text{Im}(z(n))|^p)^{\frac{1}{p}} \end{aligned} \quad (15)$$

for large  $p$ .

Convexity is preserved in this approximation [7]. In actual implementation, we can minimize the cost

$$J(\mathbf{w}) = E|\text{Re}(z(n))|^p + E|\text{Im}(z(n))|^p \quad (16)$$

to simplify computation. Substituting (14) in (16) and taking the gradient with respect to  $\mathbf{v}^*$ , we obtain

$$\begin{aligned} \mathbf{G} = \nabla_{\mathbf{v}^*} J(\mathbf{v}) &= E \left\{ p \mathbf{U}^H \mathbf{y}(n) \left( |\text{Re}(z(n))|^{p-2} \text{Re}(z(n)) \right. \right. \\ &\quad \left. \left. - j |\text{Im}(z(n))|^{p-2} \text{Im}(z(n)) \right) \right\}. \end{aligned} \quad (17)$$

The received data is processed in a burst of  $N$  symbols. A recursive method based on the gradient descent is used to obtain the spatial-temporal equalizer coefficients. The algorithm is given by:

$$\mathbf{v}^{(k+1)} = \mathbf{v}^{(k)} - \mu \hat{\mathbf{G}}^{(k)} \quad (18)$$

where  $\mathbf{v}^{(k)}$  denotes the vector  $\mathbf{v}$  at the  $k$ -th recursion,  $\mu$  is a small step size and  $\hat{\mathbf{G}}^{(k)}$  is an estimate of the gradient (17) at the  $k$ -th recursion. This estimate is obtained by averaging over the burst:

$$\begin{aligned} \hat{\mathbf{G}}^{(k)} &= \frac{1}{N} \sum_{n=1}^N \left\{ p \mathbf{U}^H \mathbf{y}(n) \left( |\text{Re}(z^{(k)}(n))|^{p-2} \right. \right. \\ &\quad \left. \left. \text{Re}(z^{(k)}(n)) - j |\text{Im}(z^{(k)}(n))|^{p-2} \text{Im}(z^{(k)}(n)) \right) \right\}. \end{aligned} \quad (19)$$

The algorithm is initialized with  $\mathbf{v}^{(0)} = \mathbf{0}$ . Such initialization is equivalent to setting the equalizer with  $\hat{\mathbf{R}}^+ \hat{\mathbf{p}}$  (i.e. the particular LS solution in (14)). Then  $\mathbf{w}^{(k)} = \hat{\mathbf{R}}^+ \hat{\mathbf{p}} + \mathbf{U}\mathbf{v}^{(k)}$ .

#### 4.4. Simulation Results

In this section, we shall provide some simulation results on the performance of the proposed semi-blind algorithm. Three users' signals ( $K = 3$ ) are impinging

on a receiver with four sensors ( $M = 4$ ). The first user is the desired signal and the other 2 users are interferers from other co-channel cells. We shall assume that the SNR of the desired signal at the receiver is 30 dB. The signal-to-interference ratio (SIR) is 3 dB in our simulations. The signals go through their respective channels which are modeled as 3 taps. This is the case when the delay spread is around 3 – 4 symbol periods. At the receiver, each sensor has an equalizer of length 6. Hence the spatial-temporal equalizer has a total of 24 coefficients.

When implementing the semi-blind algorithm (16), the choice of the exponent  $p$  has to be determined. Figure 1 shows a plot of the MSE achieved using different  $p$ 's for 16-QAM signals. The MSE is lower when a larger  $p$  is used. However, a compromise has to be struck. Using too large a  $p$  might have numerical problems in the recursion at the initial stage when the noise and ISI is severe while using too small a  $p$  does not approximate (16) well. The pure blind convex algorithm in [6] uses  $p = 12$ . We shall also use this value of  $p$  in subsequent simulations. The step size  $\mu$  for the recursive algorithm is 0.001. The performance measure is the mean square error (MSE) of the output. We shall compare the MSE among the convex semi-blind, regularized semi-blind and pure LS algorithms in the case where  $N_t < MN_w$ . The blind algorithm with tap-anchoring constraint (12) is also implemented using a recursion similar to (18) but in terms of  $\mathbf{w}$ . The blind case (which does not take into account of known symbols present in the burst) fails to converge under this scenario for both 4-QAM and 16-QAM (Fig. (2) and Fig. (3)). An AGC is used at the output for the convex semi-blind algorithm so that the comparison is meaningful. The AGC adjusts the gain by

$$A = \left( \frac{E|a_n|^2}{|z(n)|^2} \right)^{\frac{1}{2}}, \quad (20)$$

where  $a_n$  is the alphabets in the constellation and  $|z(n)|^2$  is the average over the burst. The term  $E|a_n|^2$  can be pre-computed since the constellation is known. This is, in fact, the variance of the constellation and in our simulations, we set  $E|a_n|^2 = 1$ .

Figure (2) shows the MSE vs.  $N_t$  for the case of 4-QAM signals. The MSE is that of the desired user. The burst has 150 symbols. The LS curve indicates the MSE if we are only using the training sequence to compute the equalizer coefficients. It is also an indication on the MSE before passing through the semi-blind algorithms since we initialize the algorithms using the LS solution. The regularized semi-blind algorithm is implemented as in [1]. Our convex semi-blind algorithm runs for 500 recursions. The MSE plot is obtained by

averaging over 40 runs of bursts of 150 symbols. The regularized semi-blind algorithm achieves smaller MSE in this scenario than that of the convex semi-blind algorithm.

The next simulation is on 16-QAM signals. In this case the MSE vs.  $N_t$  plot (Fig. (3)) is obtained by averaging 40 runs of bursts of 200 symbols. The convex semi-blind algorithm iterates 500 times. We can see that in this scenario, it has a smaller MSE starting from  $N_t = 12$  than the regularized semi-blind algorithm. The latter method does not perform as good as in the case of 4-QAM signals. If we can tolerate an MSE of no more than, say, 0.05, then the regularized semi-blind method will fail in this case while the convex semi-blind method is suitable for  $N_t > 16$  in a burst.

## 5. CONCLUSIONS

In this paper, a convex cost with training constraint is proposed for semi-blind adjustment of the coefficients of a spatial-temporal equalizer in general. Compared to other blind and semi-blind methods in a short burst communication scenario, the proposed method performs better especially with non-constant modulus signal constellations. Such type of constellation is proposed in the 3rd generation wireless standard when higher data rates are needed.

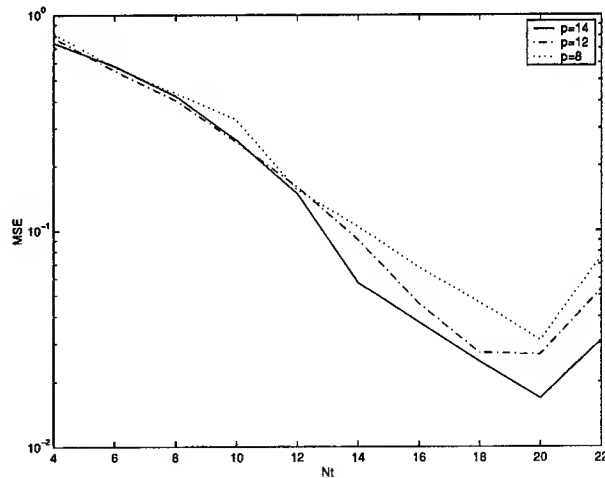


Figure 1: Plot of MSE vs.  $N_t$  for the semi-blind convex algorithm using different  $p$  ( $K = 3$ , 16-QAM signals,  $N = 200$ ).

## REFERENCES

- [1] A. Kuzminskiy, L. Féty, P. Forster, S. Mayrargue, "Regularized semi-blind estimation of spatio-temporal filter coefficients for mobile radio communications," in *Proc. GRETSI'97*, pp. 127–130, Grenoble, 1997.
- [2] D. Godard, "Self-recovering equalization and carrier tracking in two-dimensional data communication systems," in *IEEE Transactions on Communications*, vol. COM-28, pp. 1867–1875, November 1980.
- [3] Z. Ding, "On convergence analysis of fractionally spaced adaptive blind equalizers," in *IEEE Trans. on Signal Processing*, vol. 44, pp. 650–657, March 1997.
- [4] Y. Li, K. Riu and Z. Ding, "Length- and cost-dependent local minima of unconstrained blind channel equalizers," in *IEEE Trans. on Signal Processing*, vol. 44, pp. 2726–2735, November 1996.
- [5] W. A. Sethares, R. A. Kennedy and Z. Gu, "An approach to blind equalization of non-minimum phase systems," in *ICASSP*, pp. 1529–1532, 1991.
- [6] R. A. Kennedy and Z. Ding, "Blind adaptive equalizers for quadrature amplitude modulated communication systems based on convex cost functions," in *Optical Engineering*, vol. 31, pp. 1189–1199, June 1992.
- [7] S. Vembu, S. Verdú, R. A. Kennedy and W. Sethares, "Convex cost functions in blind equalization," in *IEEE Trans. on Signal Processing*, vol. 42, pp. 1952–1960, August 1994.

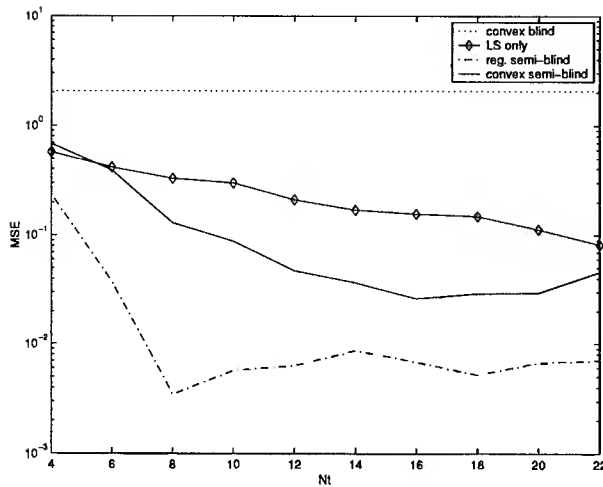


Figure 2: 4-QAM case: MSE vs.  $N_t$  for pure LS, convex blind, convex semi-blind and regularized semi-blind algorithms ( $K = 3, N = 150$ ).

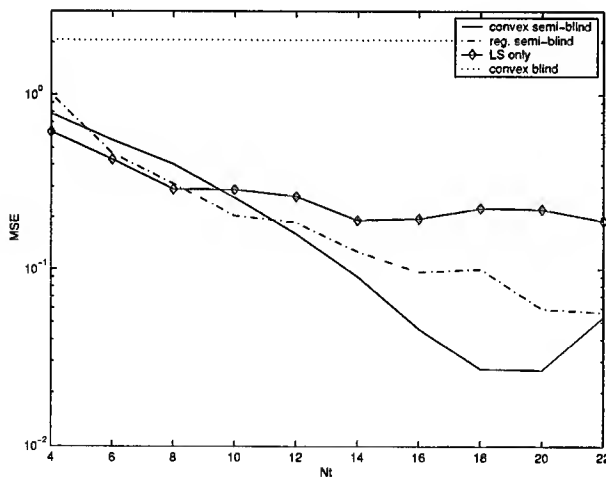


Figure 3: 16-QAM case: MSE vs.  $N_t$  for the pure LS, convex blind, convex semi-blind and regularized semi-blind algorithms ( $K = 3, N = 200$ ).

# Performance Analysis of Blind Carrier Phase Estimators for General QAM Constellations

E. Serpedin<sup>1</sup> (contact author), P. Ciblat<sup>2</sup>, G. B. Giannakis<sup>3</sup>, and P. Loubaton<sup>2</sup>

<sup>1</sup> Dept. of Electrical Engineering, Texas A&M University, College Station, TX 77843-3128, Tel.: (979) 458 2287

Fax: (979) 862 4630 email: serpedin@ee.tamu.edu

<sup>2</sup> Université de Marne-la-Vallée, Laboratoire "Systèmes de Communication", 5 Bd. Descartes, 77454 Marne-la-Vallée cedex 2, France

<sup>3</sup> Dept. of Electrical and Computer Engr., University of Minnesota, 200 Union St. SE, Minneapolis, MN 55455

**Abstract**—Large quadrature amplitude modulation (QAM) constellations are currently used in throughput efficient high speed communication applications such as digital TV. For such large signal constellations, carrier phase synchronization is a crucial problem because for efficiency reasons the carrier acquisition must often be performed blindly, without the use of training or pilot sequences. The goal of the present paper is to provide thorough performance analysis of the blind carrier phase estimators that have been proposed in the literature and to assess their relative merits.

## I. INTRODUCTION

Fast acquisition of the carrier phase is a crucial issue in high-speed communication systems that employ large QAM modulation schemes. One of the challenges associated with large QAM constellations is the blind carrier acquisition, which is often required in large and heavily loaded multipoint networks for bandwidth efficiency and little effort involved in network monitoring. It is known that for large QAM constellations, the conventional carrier tracking schemes frequently fail to converge and result in "spinning" [8], [10]. Therefore, developing computationally simple blind carrier phase estimators with guaranteed convergence and good statistical properties is well-motivated.

Recently, a number of blind carrier phase estimators have been proposed [1], [2], [3], [4], [6], [11, p. 266-277], [12], but thorough performance analysis of all these algorithms has not been performed. In order to quantify the performance of these estimators, the large sample (asymptotic) performance analysis of these phase estimators will be established and compared with the stochastic (modified) Cramér-Rao bound [11, Section 2.4]. It is shown that the seemingly different estimators [1], [2], [3], [5], [11, p. 266-277], [12], are the same, while the estimator proposed in [4] has a larger asymptotic variance than the power-law estimator [3], [6], [12]. It is also shown that by exploiting the additional samples acquired through oversampling the received continuous-time waveform does not improve the performance of the power-law estimator in [3], [6], [12]. Finally, computer simulations are presented to corroborate the theoretical developments and to compare the performance of the investigated phase estimators.

## II. PROBLEM STATEMENT

We consider the baseband QAM communication system where the received signal  $Y(n) = Y_r(n) + jY_i(n)$  is given by

$$Y(n) = e^{j\theta} X(n) + N(n), \quad (1)$$

where  $Y_r(n)$  and  $Y_i(n)$  denote the in-phase and quadrature components of  $Y(n)$ ,  $X(n)$  stands for the independent and identically distributed (i.i.d.) input QAM symbol stream,  $N(n)$  is the circularly distributed Gaussian noise, assumed to be independent of  $X(n)$ , and  $\theta$  denotes the unknown carrier phase offset. The problem of blind carrier phase estimation consists of recovering the phase error  $\theta$  only from knowledge of the received data  $Y(n)$ . Because the input QAM constellation has quadrant ( $\pi/2$ ) symmetry, it follows that it is possible to recover the unknown phase  $\theta$  only modulo a  $\pi/2$ -phase ambiguity. This ambiguity can be further eliminated through the use of appropriate coding schemes. Therefore, without any loss of generality, we can assume that the unknown phase  $\theta$  lies the interval  $(-\pi/4, \pi/4)$ . In the next section, we briefly outline the blind phase estimators [1], [2], [3], [4], [5], [11, p. 266-277], [12], and establish their exact large sample performance.

## III. BLIND CARRIER PHASE ESTIMATORS

### A. Approximate Maximum Likelihood Estimator: Fourth-Power Estimator

The maximum likelihood (ML) estimator of  $\theta$  can be theoretically derived by maximizing a stochastic likelihood function, obtained by averaging the conditional probability density function of the received data with respect to the unknown data stream  $X(n)$ . However, for high order QAM constellations, the computational complexity involved in calculating the likelihood function and more importantly the resulting nonlinear optimization problem render the ML-estimator impractical for most high-speed applications. The need for computationally simple estimators with guaranteed convergence calls for alternative (possibly suboptimal, but computationally feasible) phase estimators.

Moeneclaey and de Jonghe have shown in [12] that for any arbitrary 2-dimensional rotationally symmetric constellations (such as square or cross QAM constellations) the fourth-power (or power-law) estimator can be obtained as an approximate ML-estimator in the limit of small Signal-to-Noise Ratio ( $\text{SNR} := 10 \log E|X(n)|^2 / E|N(n)|^2$ , where  $:=$  stands for "is defined as"). The power-law estimator and its sampled version are defined as:

$$\theta := \frac{1}{4} \arg [(EX^{*4}(n)) EY^4(n)], \quad (2)$$

$$\hat{\theta} := \frac{1}{4} \arg \left[ E(X^{*4}(n)) \frac{\sum_{n=1}^N Y^4(n)}{N} \right], \quad (3)$$

where the superscript  $*$  stands for complex conjugation and the operator  $E(\cdot)$  denotes the expectation operator. The fourth-power estimator does not require any complex nonlinear optimizations, but it requires a-priori knowledge of the input constellation  $E(X^{*4}(n))$ . However, this is not a restrictive assumption since for most QAM constellations,  $EX^{*4}(n)$  is a negative real-valued number, whose effect can be easily accounted for. Using standard convergence results [9] it can be checked that asymptotically (3) is<sup>1</sup> w.p. 1 a consistent estimator ( $\hat{\theta} \rightarrow \theta$  as  $N \rightarrow \infty$ ) for any SNR range. An explanation can be obtained by observing that, in the presence of circularly and normally distributed noise  $N(n)$ , the following relation holds:

$$\frac{1}{N} \sum_{n=1}^N Y^4(n) \xrightarrow{\text{w.p.1}} EY^4(n) = e^{j4\theta} EX^4(n), \quad (4)$$

where the second equality in (4) is obtained by expanding  $EY^4(n) = E(\exp(j\theta)X(n) + N(n))^4$ , taking into account the independence between  $X(n)$  and  $N(n)$ , and  $EN^k(n) = 0$ , for any positive integer  $k$ . Hence, (3) recovers the carrier phase from the phase of the fourth-order moment of the received data.

Cartwright has proposed estimating the unknown phase  $\theta$  using a different set of fourth-order statistics [3]. Define the following fourth-order moments and cumulants:

$$\gamma := E[Y_r^4] + E[Y_i^4] - 6E[Y_r^2 Y_i^2], \quad (5)$$

$$\begin{aligned} \gamma_a &:= \text{cum}(Y_r, Y_r, Y_r, Y_i) = E[Y_r^3 Y_i] - 3E[Y_r^2]E[Y_r Y_i] \\ &= E[Y_r^3 Y_i], \end{aligned} \quad (6)$$

$$\begin{aligned} \gamma_b &:= \text{cum}(Y_r, Y_i, Y_i, Y_i) = E[Y_r Y_i^3] - 3E[Y_i^2]E[Y_r Y_i] \\ &= E[Y_r Y_i^3], \quad (E[Y_r Y_i] = 0). \end{aligned} \quad (7)$$

Cartwright's estimator is defined by:

$$\tan(4\theta) = 4 \left( \frac{\gamma_a - \gamma_b}{\gamma} \right) \Rightarrow \theta = \frac{1}{4} \text{atan} \left[ 4 \left( \frac{\gamma_a - \gamma_b}{\gamma} \right) \right]. \quad (8)$$

To verify that Cartwright's estimator is the fourth-power estimator in (2), we equate the in-phase and quadrature components of:

$$\begin{aligned} EY^4(n) &= e^{j4\theta} EX^4(n) = \cos(4\theta) EX^4(n) + j \sin(4\theta) EX^4(n) \\ EY^4(n) &= E(Y_r(n) + jY_i(n))^4 = E[Y_r^4(n) + Y_i^4(n) - 6Y_r^2(n) \\ &\quad \times Y_i^2(n) + 4jE[Y_r^3(n)Y_i(n) - Y_r(n)Y_i^3(n)] \\ &= \gamma + 4j(\gamma_a - \gamma_b). \end{aligned} \quad (9)$$

It follows that  $\gamma = \cos(4\theta) EX^4(n)$  and  $4(\gamma_a - \gamma_b) = \sin(4\theta) EX^4(n)$ , which implies the equivalence between estimators (2) and (8). Cartwright's (fourth-power) estimator requires only that  $EX^4(n) \neq 0$  and the independence between  $X(n)$  and additive circularly and normally distributed noise  $N(n)$ , and it can be applied to both square and cross-QAM constellations, as opposed to the estimator proposed in [4], which can be applied only to square-QAM constellations.

It is interesting to remark that three other phase estimators, derived using completely different arguments, are equivalent to the fourth-power estimator. An alternative robust

phase estimator with guaranteed convergence has been proposed in [2] for square-QAM constellations. Herein, the carrier acquisition problem is reduced to the blind source separation problem of the linear mixture of the in-phase and quadrature-phase components of the received signal, and a cumulant-based source separation criterion is proposed to estimate the unknown phase-offset [2]. In [1], [11, pp. 271-277], a low SNR approximation of the likelihood function, assuming PSK input constellations, is shown to have the same form as the estimator [2]. Furthermore, it is justified that this estimator can be used even for general QAM constellations [11, pp. 271-277]. By relying on Godard's quartic criterion [8], Foschini has shown an alternative derivation of this phase estimator in [5]. Next, we describe briefly the estimator proposed in [2], which relies on the observation that the in-phase and quadrature components of a square-QAM constellation are independent.

Let  $\phi$  denote an estimate of the unknown phase offset  $\theta$ , define the "rotated" output  $\tilde{Y}(n) := \exp(-j\phi) Y(n)$ , and assume that  $X(n)$  belongs to a square-QAM constellation. In the absence of noise and if  $\phi = \theta$ , then the in-phase and quadrature components of  $\tilde{Y}(n) = X(n)$  are independent. Thus, the joint cumulants of the in-phase ( $\tilde{Y}_r(n)$ ) and quadrature ( $\tilde{Y}_i(n)$ ) components of  $\tilde{Y}(n)$  are equal to zero

$$\begin{aligned} \tilde{\gamma}_a &:= \text{cum}(\tilde{Y}_r(n), \tilde{Y}_r(n), \tilde{Y}_r(n), \tilde{Y}_i(n)) = 0, \\ \tilde{\gamma}_b &:= \text{cum}(\tilde{Y}_r(n), \tilde{Y}_i(n), \tilde{Y}_i(n), \tilde{Y}_i(n)) = 0, \end{aligned} \quad (10)$$

and<sup>2</sup>  $\tilde{\gamma}_a - \tilde{\gamma}_b = 0$ . It is interesting to remark that (10) continues to hold true even in the presence of additive circularly and normally distributed noise  $N(n)$ , because the cumulants of the in-phase and quadrature components of  $N(n)$  cancel out. By taking into account (9), it follows that  $\tilde{\gamma}_a - \tilde{\gamma}_b = (E\tilde{Y}^4(n) - E\tilde{Y}^{*4}(n))/8j$ . Thus,  $\theta$  can be estimated from:

$$\begin{aligned} \theta_a &:= \arg \min_{\phi} (E\tilde{Y}^4(n) - E\tilde{Y}^{*4}(n)) \\ &= \arg \min_{\phi} (e^{-j4\phi} EY^4(n) - e^{j4\phi} EY^{*4}(n)). \end{aligned} \quad (11)$$

If we consider the polar representation  $EY^4(n) = \lambda^4 \exp(j4\theta)$ , from (11) we obtain that  $\theta_a = \arg \min_{\phi} \lambda^4 (\exp(-j4(\phi - \theta)) - \exp(j4(\phi - \theta)))$ , which implies that  $\theta_a = \theta$  modulo a  $\pi/4$ -phase ambiguity. Hence, estimator (11) is the same as the fourth-power estimator (2). By taking advantage of the sign of  $\tilde{\gamma} := (E\tilde{Y}^4(n) + E\tilde{Y}^{*4}(n))/2$  (see (5), (9)), the  $\pi/4$ -phase ambiguity inherent in (11) can be reduced to a  $\pi/2$ -phase ambiguity (since if  $\theta_a - \theta = \pi/4$  modulo  $\pi/2$ , then  $\tilde{\gamma} = -EX^4(n) \neq EX^4(n)$ ).

In practice, many communication systems utilizing QAM constellations employ also coding, which implies that the SNR available at the synchronizer will be reduced by an amount proportional to the coding gain. In order to evaluate correctly the performance of these phase estimators at all SNR levels, next we provide an exact expression for the large sample variance of the power-law estimator, which is valid for any SNR level and it is not restricted to the high SNR regime as is the case with the approximate asymptotic expression presented in [12]. The next section will show that

<sup>1</sup>The notation w.p. 1 denotes convergence with probability one.

<sup>2</sup>The reader can easily check that  $\tilde{\gamma}_a = -\tilde{\gamma}_b$ , [4].



the expression of [12] is not valid for low and medium SNRs ( $\leq 20$  dB).

**Theorem 1.** Assuming that the i.i.d. symbol stream  $X(n)$  is coming from a finite dimensional QAM-constellation and that the additive noise  $N(n)$  is circularly and normally distributed and independent of  $X(n)$ , then the estimate (3) is asymptotically normally distributed with zero mean and the asymptotic variance:

$$\lim_{N \rightarrow \infty} N(\hat{\theta} - \theta)^2 = \frac{\mu_{Y,44} - EX^8(n)}{32(EX^4(n))^2}, \quad (12)$$

with<sup>3</sup>  $\mu_{Y,40} := EY^4(n) = e^{j4\theta} EX^4(n)$ , and

$$\begin{aligned} \mu_{Y,44} := & E|X(n)|^8 + 16E|X(n)|^6 E|N(n)|^2 + 36E|X(n)|^4 \\ & \times E|N(n)|^4 + 16E|X(n)|^2 E|N(n)|^6 + E|N(n)|^8. \end{aligned} \quad (13)$$

**Proof.** Please see [13].  $\square$

The asymptotic variance (12) does not depend on the unknown phase  $\theta$ , but only on the input symbol constellation and the SNR. This confirms the conclusion drawn in [3] stating that the standard deviation of (8) appears to be constant with respect to the true value of  $\theta$ . We evaluate next the asymptotic performance of a phase estimator based on an alternative set of statistics that was proposed in [4].

#### B. HOS-Based Phase Estimator of [6]

The phase estimator [4] extracts the unknown phase information  $\theta \in (-\pi/4, \pi/4)$  using the relations:

$$\begin{aligned} \cot(2\theta) &= \frac{\gamma_a - \gamma_b}{2\gamma} \quad \text{if } \left| \frac{\gamma}{\gamma_x} \right| \geq 0.125 \Leftrightarrow \\ \theta &\in \left( -\frac{\pi}{4}, -\frac{\pi}{8} \right] \cup \left[ \frac{\pi}{8}, \frac{\pi}{4} \right), \end{aligned} \quad (14)$$

$$\begin{aligned} \tan(2\theta) &= \frac{2(\gamma_a - \gamma_b)}{\gamma_x - 4\gamma} \quad \text{if } \left| \frac{\gamma}{\gamma_x} \right| < 0.125 \Leftrightarrow \\ \theta &\in \left( -\frac{\pi}{8}, \frac{\pi}{8} \right), \end{aligned} \quad (15)$$

with  $\gamma_x := E[|X|^4] - 2\{E|X|^2\}^2$  and

$$\begin{aligned} \gamma := & \text{cum}\{Y_r(n), Y_r(n), Y_i(n), Y_i(n)\} = E[Y_r^2(n)Y_i^2(n)] \\ & - E[Y_r^2(n)]E[Y_i^2(n)] = 0.25 \sin^2(2\theta)\gamma_x. \end{aligned} \quad (16)$$

Let  $\hat{\gamma}_a$ ,  $\hat{\gamma}_b$ , and  $\hat{\gamma}$  denote sample estimates for  $\gamma_a$ ,  $\gamma_b$ , and  $\gamma$ , respectively, and define by  $\hat{\theta}_1$  and  $\hat{\theta}_2$  the sample estimates corresponding to (14) and (15), respectively. The next theorem, whose proof is deferred due to space limitations to [13], establishes the asymptotic performance of  $\hat{\theta}_1$  and  $\hat{\theta}_2$ .

**Theorem 2.** Assuming that the i.i.d. symbol stream  $X(n)$  is coming from a finite dimensional QAM-constellation and that the additive noise  $N(n)$  is circularly and normally distributed and independent of  $X(n)$ , then the estimates  $\hat{\theta}_1$  and  $\hat{\theta}_2$  are asymptotically normally distributed with zero mean and asymptotic variances:

$$\lim_{N \rightarrow \infty} N(\hat{\theta}_1 - \theta)^2 = \frac{\varrho_{11} + \cot^2(2\theta)\varrho_{22} - 2\cot(2\theta)\varrho_{12}}{\gamma_x^2},$$

<sup>3</sup> The notation  $\mu_{Y,k,l} := EY^k(n)Y^{*l}(n)$  stands for the  $(k+l)$ th-moment of  $Y(n)$ .

$$\text{if } \theta \in \left( -\frac{\pi}{4}, -\frac{\pi}{8} \right] \cup \left[ \frac{\pi}{8}, \frac{\pi}{4} \right), \quad (17)$$

$$\begin{aligned} \lim_{N \rightarrow \infty} N(\hat{\theta}_2 - \theta)^2 &= \frac{\varrho_{11} + 4\tan^2(2\theta)\varrho_{22} + 4\tan(2\theta)\varrho_{12}}{\gamma_x^2}, \\ \text{if } \theta &\in \left( -\frac{\pi}{8}, \frac{\pi}{8} \right), \end{aligned} \quad (18)$$

where:

$$\begin{aligned} \varrho_{11} := & \lim_{N \rightarrow \infty} NE[(\hat{\gamma}_a - \hat{\gamma}_b) - (\gamma_a - \gamma_b)]^2 = -\frac{|EX^4(n)|^2}{32} \\ & + \frac{\cos(8\theta)[(EX^4(n))^2 - EX^8(n)] + \mu_{Y,44}}{32}, \end{aligned} \quad (19)$$

$$\begin{aligned} \varrho_{12} := & \lim_{N \rightarrow \infty} NE\{(\hat{\gamma} - \gamma)[(\hat{\gamma}_a - \hat{\gamma}_b) - (\gamma_a - \gamma_b)]\} \\ = & \frac{-\sin(8\theta)[EX^8(n) - 2(EX^4(n))^2] + 2\text{Im}\{\mu_{Y,62}\}}{64} \\ & - \frac{4\sin(4\theta)EX^4(n)[\mu_{Y,22} - 3\mu_{Y,11}^2]}{64} \\ & - \frac{8(E|X(n)|^2 + E|N(n)|^2)\text{Im}\{\mu_{Y,51}\}}{64}, \end{aligned} \quad (20)$$

$$\begin{aligned} \varrho_{22} := & \lim_{N \rightarrow \infty} NE(\hat{\gamma} - \gamma)^2 = \frac{\cos(8\theta)EX^8(n) + 3\mu_{Y,44}}{128} \\ & - \frac{4\text{Re}\{\mu_{Y,62}\} + 48\mu_{Y,11}^4 + 6[\cos(4\theta)EX^4(n) - \mu_{Y,22}]^2}{128} \\ & - \frac{32\mu_{Y,11}^2[\cos(4\theta)EX^4(n) - 2E|Y(n)|^4]}{128} \\ & + \frac{16[\text{Re}\{\mu_{Y,51}\} - \mu_{Y,33}]\mu_{Y,11}}{128}, \end{aligned} \quad (21)$$

$\mu_{Y,44}$  is given by (13), and

$$\mu_{Y,62} := e^{j4\theta}[EX^6(n)X^{*2}(n) + 12EX^5(n)X^*(n)E|N(n)|^2 + 15EX^4(n)E|N(n)|^4], \quad (22)$$

$$\mu_{Y,51} := e^{j4\theta}[EX^5(n)X^*(n) + 5EX^4(n)E|N(n)|^2], \quad (23)$$

$$\begin{aligned} \mu_{Y,33} := & E|X(n)|^6 + 9E|X(n)|^4 E|N(n)|^2 \\ & + 9E|X(n)|^2 E|N(n)|^4 + E|N(n)|^6, \end{aligned} \quad (24)$$

$$\mu_{Y,22} := E|X(n)|^4 + 4E|X(n)|^2 E|N(n)|^2 + E|N(n)|^4, \quad (25)$$

$$\mu_{Y,11} := E|X(n)|^2 + E|N(n)|^2. \quad (26)$$

Opposed to the power-law estimator, the asymptotic performance of the Chen et al. estimator [4] depends on the phase offset  $\theta$ . As the simulation results will show (see Figure 5), the asymptotic performance of this estimator deteriorates significantly whenever the a-priori intervals (14), (15) are missed, and for any SNR it exhibits a larger variance than the power-law estimator.

#### IV. PERFORMANCE COMPARISONS

In this section, computer simulations are performed to assess the relative merits of the proposed phase estimators by comparing the theoretical (asymptotic) limits and the experimental standard deviations of the investigated estimators. Two additional estimators have been analyzed: the fractionally-sampled (FS) power-law estimator and the reduced-constellation power estimator. The FS-power estimator recovers the unknown phase offset  $\theta$  by exploiting

all the samples obtained by fractionally-sampling (oversampling) the received continuous-time waveform in the estimator (3). A raised-cosine pulse shape with roll-off factor 0.3 and an oversampling factor  $P = 3$  are assumed throughout the simulations. The reduced-constellation power estimator relies also on (3), but only the received samples that are larger in magnitude than a given threshold are processed [10, p. 1382], [6, p. 1482]. Thus, only the points closest to the four corners of the constellation are processed. The asymptotic performance of these two additional estimators can be established using the result of Theorem 1, but due to space limitations their expressions will not be presented.

In Figures 1-a and b, we have plotted the experimental and theoretical standard deviations of all these estimators versus SNR, assuming a square 256-QAM constellation,  $\theta = 15^\circ (= \pi/12)$ ,  $N = 512$  samples,  $MC = 300$  Monte-Carlo runs, and additive normally distributed noise. The threshold in the reduced-constellation power estimator has been set up so that only the received samples corresponding to the 12 points of the input 256-QAM constellation with the largest radii are processed. The solid line denotes the stochastic Cramér-Rao bound (CRB =  $1/(N \cdot \text{SNR})$ ) corresponding to the phase estimate. Figure 1 shows that the power-law estimator performs better than the Chen et al. estimator [4] at all SNR levels, but worse than the reduced-constellation power estimator at high SNRs ( $\text{SNR} \geq 20$  dB). The FS-based power estimator appears to have the worst performance. The reduced performance of the FS-power estimator is due to the increased "self-noise" generated by the residual intersymbol interference effects. For this reason, we have not pursued further the analysis of FS-based power-law estimators.

In Figure 2, we have plotted separately the theoretical and experimental standard deviations of the power-law, the reduced-constellation power-law, and the Chen et al. (15) estimators, assuming  $MC = 300$  Monte-Carlo simulation runs,  $N = 512$  samples,  $\theta = \pi/12$ , and a 256-QAM input constellation. The experimental values are well predicted by the asymptotic limits for all three estimators, but the CRB seems to be a loose bound. In Figure 3, the experimental and theoretical standard deviations of the power-law and the Chen et al. estimators are plotted versus the number of samples ( $N$ ), assuming  $\text{SNR} = 10$  dB,  $MC = 300$  Monte-Carlo runs,  $\theta = \pi/12$ . It turns out that both estimators achieve the asymptotic bound even when a reduced number of samples  $N = 250 \div 500$  are used.

In Figure 4-a, the asymptotic performance of the Chen et al. estimator (14) is analyzed, assuming  $\theta = \pi/5$ ,  $MC = 300$ , and  $N = 512$ . Figures 4-b and 5 show that the performance of the Chen et al. estimator depends on the unknown phase  $\theta$  and has a larger standard deviation than the power-law estimator for any phase offset  $\theta$  (Figure 5) and for any SNR-level (Figure 4-b). In Figure 5, the theoretical standard deviations (17) and (18) are plotted on the interval  $(-\pi/4, \pi/4)$  assuming perfect a-priori knowledge of the intervals (14), (15) where  $\theta$  lies. However, in the presence of a wrong a-priori knowledge on  $\theta$  ( $|\theta| \geq \pi/4$ ) the performance of estimator [4] deteriorates significantly.

In Figures 6 and 7, we have analyzed the performance of the power-law and the reduced-constellation power-law estimators in the case of a cross 128-QAM constellation, assum-

ing  $\theta = \pi/12$ ,  $MC = 300$ ,  $N = 4000$  samples. For such constellations, the Chen et al. estimator cannot be used since the in-phase and quadrature components of the input symbol stream are not independent. In Figures 6 and 7-a, the experimental and asymptotic standard deviations of the power-law and the reduced-constellation power-law estimators are plotted for different SNR levels. Figures 7-a,b show that the asymptotic limit predicts well the experimental results for all SNR-levels and number of samples  $N \geq 1000$ . It appears also that for cross-QAM constellations, the power-law estimator exhibits very slow convergence rate and good estimates of the phase-offset can be obtained only by using a large number of samples ( $N > 5,000$ ). Finally, Figure 8 reveals that the approximate asymptotic limit derived in [12] does not predict well the exact asymptotic limit of the power-law estimator for small and medium SNRs ( $\text{SNR} \leq 20$  dB).

## REFERENCES

- [1] A. N. D'Andrea, U. Mengali, and R. Reggiannini, "Carrier phase recovery for narrow-band polyphase shift keyed signals," *Alta Freq.*, vol. LVII, pp. 575-581, Dec. 1988.
- [2] A. Belouchrani and W. Ren, "Blind carrier phase tracking with guaranteed global convergence," *IEEE Trans. on Signal Processing*, vol. 45, no. 7, pp. 1889-1894, July 1997.
- [3] K. V. Cartwright, "Blind phase recovery in general QAM communication systems using alternative higher order statistics," *IEEE Signal Processing Letters*, vol. 6, no. 12, pp. 327-329, Dec. 1999.
- [4] L. Chen, H. Kusaka, and M. Kominami, "Blind phase recovery in QAM communication systems using higher order statistics," *IEEE Signal Processing Letters*, vol. 3, no. 3, pp. 147-149, May 1996.
- [5] G. J. Foschini, "Equalizing without altering or detecting the data," *Bell Syst. Tech. J.*, vol. 64, pp. 1885-1911, Oct. 1985.
- [6] C. Georgiades, "Blind carrier phase acquisition for QAM constellations," *IEEE Trans. on Communications*, vol. 45, no. 11, pp. 1477-1486, Nov. 1997.
- [7] F. Gini and G. B. Giannakis, "Frequency offset and symbol timing recovery in flat-fading channels: a cyclostationary approach," *IEEE Trans. on Communications*, vol. 46, no. 3, pp. 400-411, March 1998.
- [8] D. Godard, "Self recovering equalization and carrier tracking in two dimensional data communication systems," *IEEE Trans. on Communications*, vol. 28, no. 11, pp. 1867-1875, Nov. 1980.
- [9] T. Ilasan, "Nonlinear time series regression for a class of amplitude modulated cosinusoids," *Journal of Time Series Analysis*, vol. 3, no. 2, pp. 109-122, 1982.
- [10] N. Jablon, "Joint blind equalization, carrier recovery, and timing recovery for high-order QAM signal constellations," *IEEE Trans. on Signal Processing*, vol. 40, no. 6, pp. 1383-1397, June 1992.
- [11] U. Mengali and A. N. D'Andrea, *Synchronization Techniques for Digital Receivers*, Plenum, NY, 1997.
- [12] M. Moeneclaey and G. de Jonghe, "ML-oriented NDA carrier synchronization for general rotationally symmetric signal constellations," *IEEE Trans. on Communications*, vol. 42, no. 8, pp. 2531-2533, Aug. 1994.
- [13] "Proofs of Theorems 1, 2," <http://ee.tamu.edu/~serpedin>.

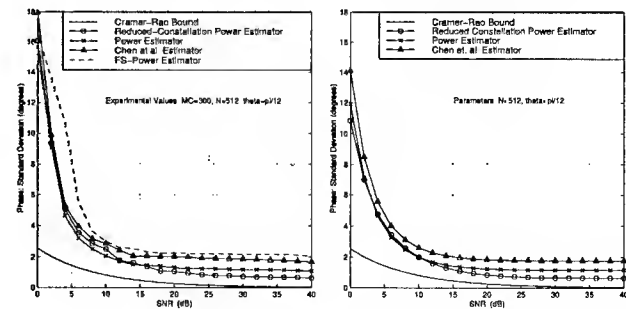


Fig. 1. Standard Deviation vs. SNR a) Experimental Values b) Asymptotic Values (256 square-QAM)

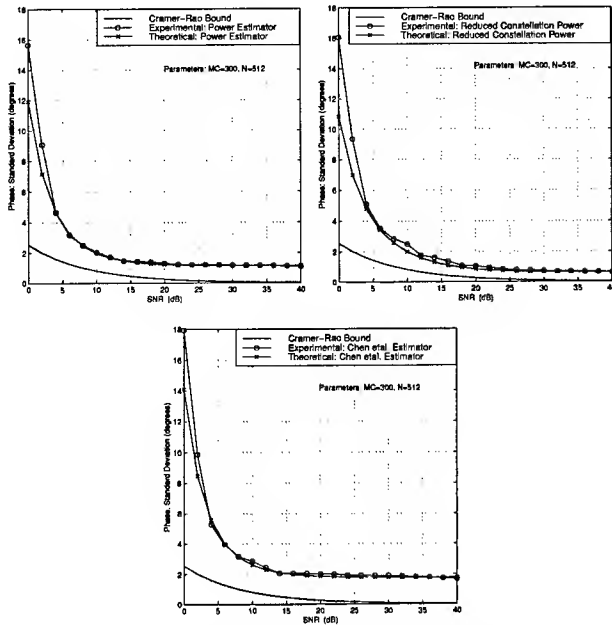


Fig. 2. Standard Deviation vs. SNR: Experimental/Theoretical Values a) Power Estimator b) Reduced-Constellation Power Estimator c) Chen et al. Estimator (256 square-QAM)

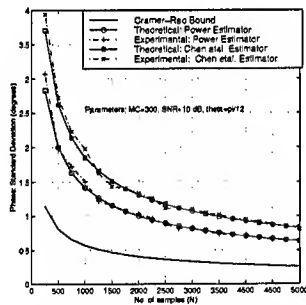


Fig. 3. Standard Deviation vs. No. of Samples: Power Estimator vs. Chen et al. Estimator (256 square-QAM)

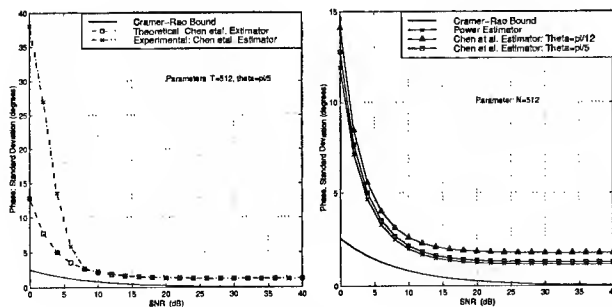


Fig. 4. Standard Deviation vs. SNR a) Chen et al. Estimator ( $\theta = \pi/5$ ) b) Asymptotic Limits (256 square-QAM)

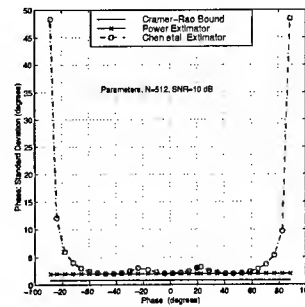


Fig. 5. Standard Deviation vs. Phase offset: Asymptotic Limit (256 square-QAM)

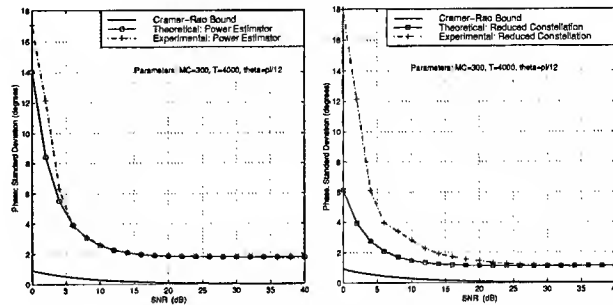


Fig. 6. Standard Deviation vs. SNR a) Power Estimator b) Reduced-Constellation Power Estimator (128 cross-QAM)

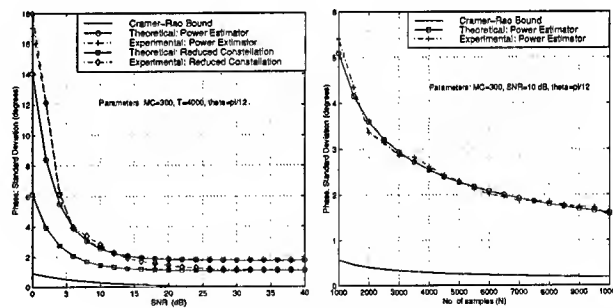


Fig. 7. Standard Deviation vs. SNR/Data: a) Reduced-Constellation Power-Law and Power-Law Estimators b) Power Estimator (128 cross-QAM)

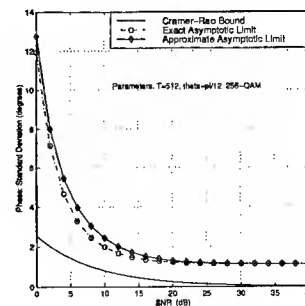


Fig. 8. Standard Deviation vs. SNR: Exact and Approximate Asymptotic Limits (256 square-QAM)

# UNBIASED PARAMETER ESTIMATION FOR THE IDENTIFICATION OF BILINEAR SYSTEMS

*Souad MEDDEB, Jean Yves TOURNERET and Francis CASTANIE*

ENSEEIH /TESA, 2 Rue Camichel, 31071 , Toulouse, France

e-mail: meddeb@len7.enseeiht.fr

## ABSTRACT

This paper addresses the problem of time-invarying (TIV) bilinear system identification. The input-output relation of a TIV bilinear system is expressed as a time-varying recursive equation. Such formulation allows us to estimate the unknown bilinear system parameters using a modified least-squares (MLS) algorithm. The MLS method provides unbiased estimates of the unknown bilinear parameters. Several simulations illustrate the MLS estimator performance.

## 1. INTRODUCTION

Linear models have found a variety of applications in many areas such as speech processing, image processing and communications. These models include parametric Autoregressive (AR), Moving Average (MA) or Autoregressive Moving Average (ARMA) models. The use of these parametric models can be motivated by the following property: for any real-valued stationary process  $y(n)$  with continuous spectral density  $S(f)$ , it is possible to find an ARMA process whose spectral density is arbitrarily close to  $S(f)$  ([2], p. 130). However, these models fail to identify many systems which are inherently nonlinear.

Bilinear model has been used successfully to approximate a large class of nonlinear systems [5][7]. Its ability to represent many nonlinearities efficiently and with a relatively small number of parameters is owing to its feedback structure [5]. Other properties motivating the use of bilinear systems are also discussed in [4]. The problem of estimating bilinear system parameters using measurements of the system input and output signals has received much attention in the literature [3][6]. Recursive estimation algorithms including the recursive least squares algorithm (RLS) or the extended least squares algorithm (ELS) have been studied in [3]. The main advantage of the RLS algorithm is its simplicity because of the linearity in the parameters. However, the algorithm provides biased estimates. Simulations presented in [3] have shown that the ELS algorithm

outperforms the RLS algorithm in terms of bias. However, no theoretical study was provided because of the non-linear estimation problem and the difficult computation required. Hence various methods have been devised to obtain unbiased estimators from linear estimation problems. Some of these methods are based on modifying the least squares estimator by subtracting the bias from the estimates [8]. This paper studies the modified least squares (MLS) algorithm for the identification of bilinear systems. The MLS algorithm yields unbiased parameter estimates and lower computational cost than the ELS algorithm.

The paper is organized as follows. Section II presents the problem. Section III studies the recursive MLS algorithm for the bilinear system identification problem. Simulation results and conclusions are reported in section IV.

## 2. PROBLEM FORMULATION

The output  $x(t)$  of a bilinear system driven by the input sequence  $u(t)$  can be defined by the following recursive equation :

$$x(t) = \sum_{i=1}^p a_i x(t-i) + \sum_{i=1}^p b_i u(t-i) + \sum_{i=1}^p \sum_{j=1}^p c_{i,j} u(t-j) x(t-i) \quad (1)$$

where  $a_i, b_i, c_{ij}$  are the unknown bilinear system parameters and  $t = 1, \dots, N$ . A noisy version of  $x(t)$  denoted

$$y(t) = x(t) + e(t) \quad (2)$$

is observed (see fig. 1). In eq. (2),  $e(t)$  is a stationary white Gaussian noise with zero mean and variance

$$E[e(t)e(s)] = \sigma^2 \delta_{t,s}$$

where  $\delta_{t,s}$  is the kronecker symbol. Eq.'s (1) and (2) show that the observed process  $y(t)$  satisfies the follow-

ing time-varying (TV) model :

$$y(t) = a_0(t) + \sum_{i=1}^p a_i(t)y(t-i) + \epsilon(t) \quad (3)$$

where the TV parameters are

$$\begin{aligned} a_0(t) &= \sum_{i=1}^p b_i u(t-i) \\ a_i(t) &= a_i + \sum_{j=1}^p c_{i,j} u(t-j), i = 1, \dots, p \end{aligned} \quad (4)$$

In eq. (3),  $\epsilon(t)$  is a colored noise sequence defined by:

$$\epsilon(t) = e(t) - \sum_{i=1}^p a_i(t)e(t-i), t = 1, \dots, N$$

Model (3) is similar to the TV ARMA model studied in [1] for the identification of non-stationary signals embedded in noise. Indeed,  $a_i(t)$  can be viewed as a linear combination of functions  $f_j(t)$  as follows:

$$a_i(t) = \sum_{j=0}^p a_{ij} f_j(t), i = 0, \dots, p \quad (5)$$

with

$$\begin{aligned} a_{00} &= 0, a_{0j} = b_j, j = 1, \dots, p \\ a_{i0} &= a_i, a_{ij} = c_{ij}, j = 1, \dots, p \\ f_0(t) &= 1, f_j(t) = u(t-j), j = 1, \dots, p \end{aligned} \quad (6)$$

Eq. (5) is similar to the decomposition of the time-varying AR parameters onto a set of basis time functions studied in [1]. This paper proposes to estimate the unknown bilinear system parameters from the input and output samples  $u(t)$  and  $y(t)$  for  $t = 1, \dots, N$  using the modified least squares (MLS) algorithm [1], [8].

### 3. LEAST-SQUARES ESTIMATORS

Denote  $\theta^T = (b^T, \theta_1^T)$  the bilinear system parameter vector with  $b^T = (b_1, \dots, b_p)$  and

$$\theta_1^T = (a_1, c_{1,1}, c_{1,2}, \dots, c_{1,p}, a_2, c_{2,1}, \dots, a_p, \dots, c_{p,p}) \quad (7)$$

Eq. (3) can be written in matrix form as follows:

$$y(t) = y_{t-1}^T \theta + \epsilon(t), \quad t = 1, \dots, N \quad (8)$$

where

$$\begin{aligned} y_{t-1}^T &= (u(t-1), u(t-2), \dots, u(t-p), \\ &\quad y(t-1), y(t-1)u(t-1), \dots, y(t-1)u(t-p), \\ &\quad \dots, \\ &\quad y(t-p), y(t-p)u(t-1), \dots, y(t-p)u(t-p)) \end{aligned}$$

### 3.1. The Conventional LS Algorithm

The conventional least squares (LS) estimator of  $\theta$  denoted  $\hat{\theta}_N$ , is defined by

$$\hat{\theta}_N = \arg \min_{\theta} J_1(\theta) \quad (9)$$

where  $J_1(\theta) = \sum_{t=1}^N \epsilon^2(t)$ . Since  $J_1(\theta)$  is linear w.r.t.  $\theta$ , an analytical solution for  $\theta$  can be derived:

$$\hat{\theta}_N = \left( \sum_{t=1}^N y_{t-1} y_{t-1}^T \right)^{-1} \sum_{t=1}^N y_{t-1} y(t)$$

The white noise sequence  $e(t)$  being zero-mean and decorrelated with  $x(t)$ ,  $\lim_{N \rightarrow \infty} \hat{\theta}_N$  can be expressed as a function of the true parameter vector as follows :

$$\lim_{N \rightarrow \infty} \hat{\theta}_N = \theta - \sigma^2 \lim_{N \rightarrow \infty} P_N \sum_{t=1}^N \begin{pmatrix} 0_{p,p} & 0 \\ U_t^T & \\ & \ddots \\ 0 & U_t \end{pmatrix} \theta_1 \quad (10)$$

where  $0_{p,p}$  is the  $p \times p$  zero matrix,

$$P_N = \left( \sum_{t=1}^N y_{t-1} y_{t-1}^T \right)^{-1}$$

and

$$U_t = (1, u(t-1), \dots, u(t-p))^T (1, u(t-1), \dots, u(t-p)).$$

Eq. (10) shows that the LS estimator of  $\theta$  is generally asymptotically biased.

### 3.2. The Extended LS Algorithm

The Extended Least Squares (ELS) algorithm has shown interesting properties for pseudo-linear regression models such as (8) [3]. This algorithm can be summarized as follows:

$$\begin{aligned} Q_N &= 1 + y_N^T P_N y_N, \\ P_{N+1} &= P_N - P_N y_N Q_N^{-1} y_N^T P_N, \\ \hat{\theta}_{N+1} &= \hat{\theta}_N + P_N y_N Q_N^{-1} (y_{N+1} - y_N^T \hat{\theta}_N), \\ \bar{y}(N+1) &= y_N^T \hat{\theta}_{N+1}, \end{aligned}$$

$$y_{N+1}^T = (u(N), \dots, \bar{y}(N), \dots, \bar{y}(N+1-p)u(N+1-p)).$$

It is well known that the ELS algorithm provides unbiased estimates. However, it suffers from stability problems [3]. Next section studies another unbiased estimator known as Modified Least Squares (MLS) estimator.

### 3.3. The Modified LS Estimator

The MLS estimator also denoted bias-compensated least squares estimator is defined as follows [8]:

$$\tilde{\theta}_N = \hat{\theta}_N + \sigma^2 P_N V_N \tilde{\theta}_{1,N-1} \quad (11)$$

with

$$V_N = \sum_{t=1}^N \begin{pmatrix} 0_{p,p} & 0 \\ U_t & \\ & \ddots \\ 0 & U_t \end{pmatrix} \quad (12)$$

The MLS estimator defined by eq. (11) is clearly asymptotically unbiased. However, this estimator requires to compute the sum of  $N$  matrices of size  $(p^2 + 2p) \times (p^2 + p)$ . In order to avoid such computation, we assume in the following that the input sequence  $u(t)$  is a sequence of mutually independent and identically distributed (i.i.d) random variables with zero-mean and variance  $\sigma_u^2 = 1$ . In this case, the following result can be obtained:

$$\lim_{N \rightarrow \infty} \frac{1}{N} V_N = \begin{pmatrix} 0_{p,p} & 0 \\ I_{p+1} & \\ & \ddots \\ 0 & I_{p+1} \end{pmatrix} \triangleq V \quad (13)$$

The following biased compensated LS estimator can then be defined:

$$\tilde{\theta}_N = \hat{\theta}_N + \sigma^2 N P_N V \tilde{\theta}_{1,N-1} \quad (14)$$

Eq. (14) explicetly depends on the noise variance  $\sigma^2$ . Next section studies a recursive algorithm for the joint estimation of  $\sigma^2$  and  $\theta$  as in [8].

### 4. NOISE VARIANCE ESTIMATION FOR THE MLS ALGORITHM

Denote  $\xi_t(N)$  the residual at time  $N$  and  $R_N$  the sum of residual squares:

$$\xi_t(N) = y(t) - y_{t-1}^T \hat{\theta}_N \quad (15)$$

$$R_N = \sum_{t=1}^N \xi_t^2(N) \quad (16)$$

Eq. (8) shows that the residual  $\xi_t(N)$  can be written

$$\xi_t(N) = y_{t-1}^T (\theta - \hat{\theta}_N) + e(t) - e_{t-1}^T \theta_1$$

where

$$\begin{aligned} e_{t-1}^T &= (e(t-1), e(t-1)u(t-1), \dots, e(t-1)u(t-p), \\ &\dots, \\ &e(t-p), e(t-p)u(t-1), \dots, e(t-p)u(t-p)) \end{aligned}$$

It is well known that  $\hat{\theta}_N$  satisfies the normal equations (obtained by differentiating  $J_1(\theta)$  with respect to  $\theta$ ) [8]:

$$\sum_{t=1}^N y_{t-1} \xi_t(N) = 0$$

Consequently

$$E \left[ \frac{1}{N} R_N \right] = E \left\{ \frac{1}{N} \sum_{t=1}^N (e(t) - e_{t-1}^T \theta_1) \xi_t(N) \right\} \quad (17)$$

hence

$$E \left[ \frac{1}{N} R_N \right] = \sigma^2 + \sigma^2 E \left[ \hat{\theta}_N^T \right] V \theta_1$$

By replacing the expectation  $E \left[ \frac{1}{N} R_N \right]$  and  $E \left[ \hat{\theta}_N^T \right]$  by their instantaneous values, an estimator of the noise variance can be defined:

$$\hat{\sigma}_N^2 = \frac{1}{N} \frac{R_N}{1 + \hat{\theta}_N^T V \tilde{\theta}_{1,N-1}} \quad (18)$$

The MLS algorithm for the joint estimation of the noise variance  $\sigma^2$  and the bilinear system parameter vector  $\theta$  is then based on the following recursive equations:

$$Q_N = 1 + y_N^T P_N y_N, \quad (19a)$$

$$P_{N+1} = P_N - P_N y_N Q_N^{-1} y_N^T P_N, \quad (19b)$$

$$\hat{\theta}_{N+1} = \hat{\theta}_N + P_N y_N Q_N^{-1} (y_{N+1} - y_N^T \hat{\theta}_N), \quad (19c)$$

$$R_{N+1} = R_N + \xi_{N+1}^2 (N+1) Q_N^{-1}, \quad (19d)$$

$$\hat{\sigma}_{N+1}^2 = \frac{1}{(N+1)} \frac{R_{N+1}}{1 + \hat{\theta}_N^T V \tilde{\theta}_{1,N-1}}, \quad (19e)$$

$$\tilde{\theta}_{N+1} = \hat{\theta}_{N+1} + (N+1) \hat{\sigma}_{N+1}^2 P_{N+1} V \tilde{\theta}_{1,N-1} \quad (19f)$$

Note that eq.'s (19a), (19b) and (19c) are the classical RLS equations [3]. Eq.'s (19d), (19e) and (19f) ensure that the bilinear system parameter estimates are asymptotically unbiased. It is interesting to note that the MLS algorithm does not require any matrix inversion.

### 5. SIMULATION RESULTS

Many simulations have been performed to illustrate the previous theoretical results. For this experiment, consider the following second-order bilinear system [3]

$$\begin{aligned} x(t) &= 1.5x(t-1) - 0.7x(t-2) + u(t-1) \\ &\quad + 0.5u(t-2) + 0.12x(t-1)u(t-1) \end{aligned}$$

The observed driving sequence  $u(t)$  is white Gaussian with variance 1. The bilinear signal  $x(t)$  is contaminated by white Gaussian noise with signal-to-noise ratios (SNR's) ranging from 5 to 40dB. The algorithm is

initialized with  $\theta = 0$  and  $P_N = 1/\delta I$  where  $\delta \ll 1$ . Fig. 2 shows the convergence of the noise variance estimate to its true value ( $SNR = 5$  dB or equivalently  $\sigma^2 = 7.28$ ) from 10 Monte-Carlo simulations. The mean square errors (MSE's) of the bilinear system estimates using RLS, ELS and MLS algorithms computed from 10 Monte-Carlo simulations are depicted in fig. 3 as a function of the SNR for  $N = 4000$ . The MLS estimator clearly outperforms the usual RLS estimator in terms of MSE. Fig. 3 also shows that the MLS estimator outperforms the ELS estimator for low SNR's. Tables 2 and 3 show the bias of RLS, ELS and MLS estimates for two values of SNR. As expected, the MLS estimator outperforms the usual RLS estimator in terms of bias. The MLS and ELS algorithms perform very similarly in term of bias.

## 6. APPLICATION : NON LINEAR SATELLITE CHANNEL IDENTIFICATION

Several non linear techniques have been proposed for modeling non linear channels with memory. These techniques include Volterra series, wavelet networks and neural networks [11]. The use of Volterra series to model satellite channels was motivated in [9] and [10]. These Volterra models suffer from the number of parameters that increases exponentially with the memory and nonlinearity order. It is well known that the bilinear model can be decomposed in a Volterra series with a reduced number of parameters [4]. Consequently, this paper propose 1) to model the non linear satellite channel using the bilinear model and 2) to identify such non linear model using the LS procedures described in previous sections. A simplified satellite channel consists of two earth stations connected by a satellite repeater as depicted in fig. 4 (see [11] for more details including channel characteristics). As an example, Fig. 5 shows the normalized prediction error between the outputs of the noisy simplified satellite channel and the corresponding bilinear system computed using MLS algorithm.

## 7. CONCLUSION

The new contribution of this paper is to derive a modified least squares algorithm, from the theory of linear time-varying models for the identification of time invarying bilinear models. A recursive version of the modified least squares algorithm is derived as well. The algorithm provides estimates of the noise variance and bilinear model parameters. Bilinear MLS parameter estimates are shown to be asymptotically unbiased. The MLS estimator performance is compared to that of the

RLS and ELS estimators. The MLS estimator is finally applied to the identification of the non linear satellite channels.

## 8. REFERENCES

- [1] G. Alengrin, M. Barlaud and J. Menez, "Unbiased Parameter Estimation of Nonstationary Signals in Noise," IEEE trans. on ASSP, vol. 34, n°5, pp. 1319-1322, oct. 1986.
- [2] P. J. Brockwell and R.A. Davis, Time Series: Theory and Methods, Springer Verlag, 1990.
- [3] F. Fnaiech and L. Ljung, "Recursive Identification of Bilinear Systems," Int. J. Control, Vol. 45, No. 2, pp. 453-470, 1987.
- [4] D. Guégan, "Série Chronologiques Non Linéaires à Temps Discret", Statistique Mathématique et Probabilité, Economica.
- [5] V. John Mathews, "Adaptive Polynomial Filters," IEEE SP Magazine, pp. 10-26, July 1991.
- [6] S. Meddeb, J. Y. Tournet and F. Castanié, "Identification of Bilinear Systems Using Bayesian Inference", Proc. of ICASSP, pp. 1609-1612, Seattle, USA, May 12-15, 1998.
- [7] R. R. Mohler and W. J. Kolodziej, "An over view of bilinear system theory and applications," IEEE Transaction on Systems, Man, and Cybernetics, Vol. SMC-10, pp. 683-688, 1982.
- [8] S. Sagara and K. Wada, "On-line modified least-squares parameter estimation of linear discrete dynamic systems," Int. Jour. of Cont., Vol. 25, no. 3, pp. 329-343, 1977.
- [9] S. Benedetto, E. Biglieri and R. Daffara, "Modelling and performance evaluation of nonlinear satellite links - A Volterra series approach," IEEE Trans. AES, vol. 15, pp. 494-506, July 1979.
- [10] S. Meddeb and J. Y. Tournet, "Identification of Non-linear Satellite mobile channels using Volterra Filters," in proc. EUSIPCO, Tampere (Finland), septembre, 2000.
- [11] M. Ibnkahla, N. J. Bershad, J. Sombrin and F. Castanié, "Neural networks modelling and identification of nonlinear channels with memory: Algorithms, applications and analytic models," IEEE Trans. SP, vol. 46, no. 5, May 1998.

Estimators	RLS	ELS	MLS
$b_1$	-0.0057	0.0087	-0.0023
$b_2$	0.6487	-0.0236	-0.0145
$a_1$	-0.6614	-0.0034	0.0119
$a_2$	0.5958	-0.0004	-0.0096
$c_{1,1}$	-0.0261	-0.0034	-0.0023
$c_{1,2}$	0.0232	-0.004	-0.0026
$c_{2,1}$	0.0201	0.0039	0.0035
$c_{2,2}$	0.0491	0.0056	-0.0016

Table 1: Bias of RLS, ELS and MLS Estimates (SNR=10dB).

Estimators	RLS	ELS	MLS
$b_1$	-0.0359	-0.0397	-0.0296
$b_2$	-0.9624	0.0421	0.0300
$a_1$	-0.9443	-0.0295	-0.0120
$a_2$	0.8048	0.0204	0.0108
$c_{1,1}$	-0.0645	0.0393	0.0043
$c_{1,2}$	0.0335	0.0045	-0.0047
$c_{2,1}$	0.0311	-0.0522	-0.0101
$c_{2,2}$	0.0640	0.0070	0.0152

Table 2: Bias of RLS, ELS and MLS Estimates (SNR=5dB).

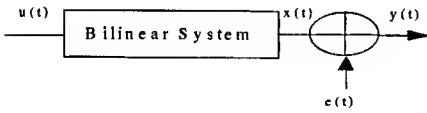


Fig. 1: Noisy Bilinear System.

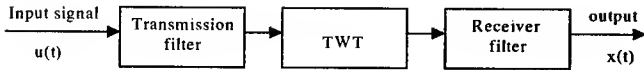


Fig. 4: Simplified satellite channel.

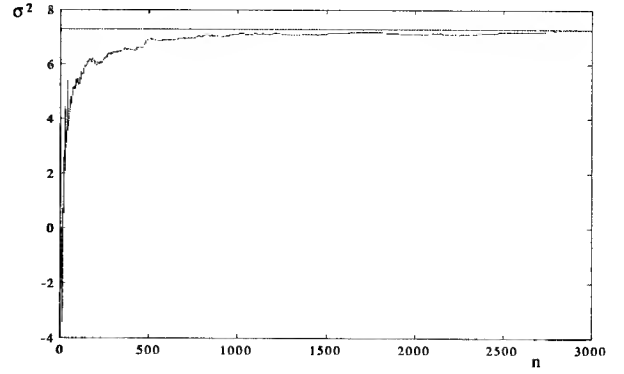


Fig. 2: Estimation of the noise variance  $\sigma^2$ .

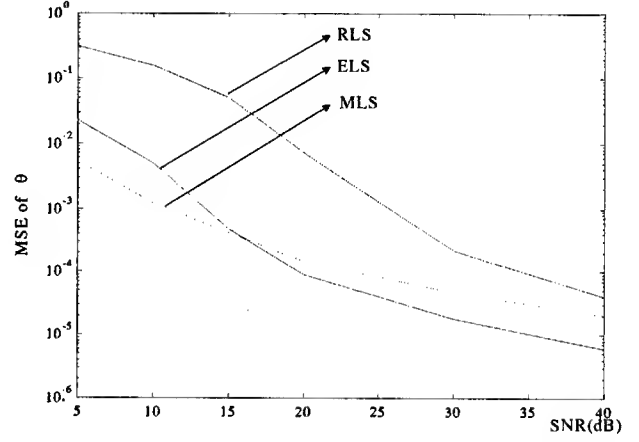


Fig. 3: MSE of the RLS, ELS and MLS estimates.

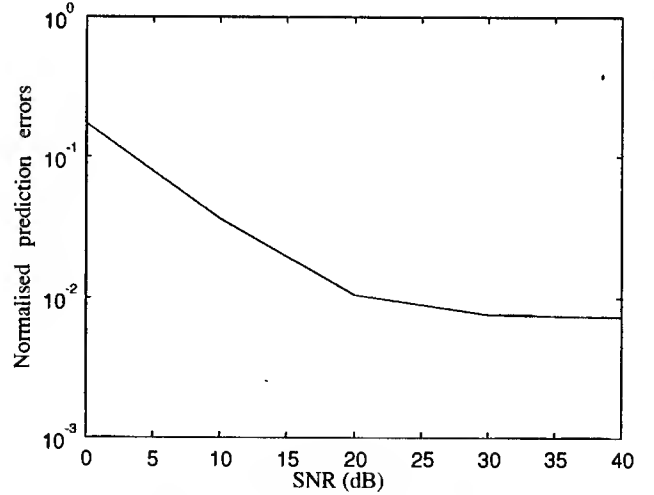


Fig. 5: Normalized prediction errors.



# BLIND IDENTIFICATION OF LINEAR-QUADRATIC CHANNELS WITH USUAL COMMUNICATION INPUTS

Nicolas PETROCHIOS<sup>1,2</sup>

(1) CAS, Dept EE, Delft Univ. of Technology  
Mekelweg 4, 2628 CD Delft, The Netherlands  
petro@ieee.org

Pierre COMON<sup>2</sup>

(2) I3S, Algorithmes-Euclide-B  
2000 route des Lucioles, BP 121  
F-06903 Sophia-Antipolis cedex, France  
comon@unice.fr

## ABSTRACT

This article presents a method to blindly identify linear quadratic channels (LQC). The method is designed for the single-input/single-output (SISO) case with white inputs with specific distributions (as those usually found in digital communications). Using High-Order Statistics (HOS) of the input, the method is able to match the third-order moments with the LQC model, yielding an original simple relation. Several simulations are performed and show a fair accuracy given sufficiently long observation records.

## 1. INTRODUCTION

Nonlinear systems provide a better approximation to real life channels, and many examples of nonlinearities can be found in nonlinear control systems [5], hydrodynamics [4], satellite communication systems [1], or underwater acoustics, among others. Blind methods are attractive when the input is unknown, and to avoid the reduction of the information rate caused by the insertion of training sequences.

Blind identification of Volterra systems has been already widely studied in the past. For instance, in [7], the authors derive the cumulant-matching equations, allowing to blindly identify a pure real quadratic system, with i.i.d. inputs of unknown distribution. Next in [2], P.Bondon goes much further, and derives identifiability conditions, when two input sequences are observed, one Gaussian and one non Gaussian.

In this paper, we focus our attention on linear-quadratic systems, with specific discrete inputs, encountered in  $n$ -PSK and QAM digital modulations. So this contribution differs from the previous ones in two respects: the system is not purely quadratic, and

This work was partly supported by ENS Lyon, ENSEA, TU-Delft, and the RNRT project "Paestum". The first author thanks A. Trindade, A. Heldring, S. Halford, and A. Elmilady for their moral support, E. Serpedin for useful discussions, and G. Giannakis for having attracted his attention on the non-linear blind identification problem.

the inputs are imposed to be discrete and of known distribution. The scope is thus less general.

## 2. MODEL FORMULATION

The problem is modeled here by the parameterization of the channel and by the statistics of the inputs.

### 2.1. Volterra kernel model

The model is described by the noisy output of a nonlinear system moving average Volterra model (which can be of any order). Sampling at a rate  $T_S$  and restricting to the Linear-Quadratic case, the channel can be modeled as:

$$y(n) = \sum_{l=0}^{L_1} h_1(l) x(n-l) + v(n) + \sum_{i,j=0}^{L_2} h_2(i,j) x(n-i) x(n-j) \quad (1)$$

where  $x(n)$  is the input signal,  $v(n)$  denotes the additive noise, and  $h_n$  is called the  $n^{\text{th}}$ -order Volterra non-linear operator (here, we only have the linear and the quadratic term:  $h_1(l_1)$  and  $h_2(l_1, l_2)$ ). Without loss of generality, we consider that  $h_n$  is symmetric in its arguments [6, pp.80-81].

### 2.2. Usual communication inputs

For the sake of convenience, denote:

$$\epsilon_{ab} \stackrel{\text{def}}{=} E[x^a x^{b*}]$$

In this article, we consider inputs commonly used in digital communications, sharing the high-order properties:

$$\epsilon_{21} = \epsilon_{31} = \epsilon_{32} = \epsilon_{41} = \epsilon_{42} = 0 \quad (2)$$

Among these inputs, two groups have been identified (see [9] and [3]):

- Distributions that are symmetric about both axes in the complex plane:  $p(z) = f(\Re\{z\}) \cdot g(\Im\{z\})$ . Corresponding random variables can be rewritten as  $z = \varepsilon + j\varepsilon'$ , where  $\varepsilon$  and  $\varepsilon'$  are real, independent, and symmetrically distributed, and  $j^2 \stackrel{\text{def}}{=} -1$ . QAM constellations, in digital communications, belong to this class.
- Discrete distributions that are invariant by a rotation of an angle of the form  $\frac{2\pi}{K}$ , ( $K \in \mathbb{N}$ ). QPSK, double QPSK, and any  $n$ -PSK are included in this class as soon that  $n \geq 4$ .

### 3. CHANNEL IDENTIFICATION

First, the basis of the identification process is presented, then the algorithms are derived, and a proof of uniqueness is eventually given.

#### 3.1. Moment-matching relations

Consider the following assumptions:

- (AS1) The channel is Linear-Quadratic of finite known length.
- (AS2) The input is stationary independent identically distributed (i.i.d.), and must comply with the properties (2);  $\sigma_x^2 = \epsilon_{11}$  and  $\mu_{4x} = \epsilon_{22}$  are also assumed to be known.
- (AS3) The noise is signal-independent white Gaussian.

Let us now define the complex bicomrelation as:

$$C_{12y}(l, k) \stackrel{\text{def}}{=} E \{ y^*(n) y(n+l) y(n+k) \} \quad (3)$$

Under assumptions (AS1-AS3), the bicomrelation of the output (3) and the channel model (1) should match, which gives the following relations:

$$C_{12y}(l, k) = \sum_{i,j} h_1(i+l) h_1(j+k) \hat{h}_2^*(i, j) \quad (4)$$

with  $(l, k) \in [-L_2, L_1] \times [-L_2, L_1]$ , and where  $\hat{h}_2^*(i, j) \stackrel{\text{def}}{=} [2\epsilon_{11}^2 + \delta(i-j)(\epsilon_{22} - 2\epsilon_{11}^2)] \cdot h_2^*(i, j)$ . The Z-transform of  $C_{12y}(l, k)$  gives in the  $(Z_1, Z_2)$  domain:

$$S_{12y}(Z_1, Z_2) = H_1(Z_1) H_1(Z_2) \hat{H}_2^* \left( \frac{1}{Z_1^*}, \frac{1}{Z_2^*} \right) \quad (5)$$

Equations (4) or (5) form the core of the algorithms subsequently proposed. By stacking the elements of  $C_{12y}(l, k)$  in a matrix  $\mathbf{C12Y}$  as:

$$\mathbf{C12Y} \stackrel{\text{def}}{=} \begin{bmatrix} C_{12y}(-L_2, -L_2) & \cdots & C_{12y}(-L_2, L_1) \\ \vdots & & \vdots \\ C_{12y}(L_1, -L_2) & \cdots & C_{12y}(L_1, L_1) \end{bmatrix},$$

we get the matrix formulation:

$$\mathbf{C12Y} = \mathbf{A}^T \cdot \mathbf{B} \cdot \mathbf{A} \quad (6)$$

where  $\mathbf{A}$  is the  $(L_2 + 1) \times (L_1 + L_2 + 1)$  Upper Triangular Band (UTB) Töplitz matrix containing  $\mathbf{h}_1 = [h(0) \dots h(L_1)]$  in the first row and zeros elsewhere:

$$\mathbf{A} \stackrel{\text{def}}{=} \begin{bmatrix} \boxed{\mathbf{h}_1} & \mathbf{0} \\ & \ddots & \ddots \\ \mathbf{0} & \boxed{\mathbf{h}_1} \end{bmatrix}$$

while  $\mathbf{B}$  is symmetric complex and contains the values of the kernel  $\hat{h}_2^*$ :

$$\mathbf{B} \stackrel{\text{def}}{=} \begin{bmatrix} \hat{h}_2^*(L_2, L_2) & \cdots & \hat{h}_2^*(0, L_2) \\ \vdots & & \vdots \\ \hat{h}_2^*(L_2, 0) & \cdots & \hat{h}_2^*(0, 0) \end{bmatrix}$$

We propose to identify the channel coefficients by using either relation (5), or (6) with the estimate  $\widehat{C_{12y}}(l, k) \stackrel{\text{def}}{=} \frac{1}{N} \sum_{n=1}^N y^*(n) y(n+l) y(n+k)$ .

One can notice that  $\mathbf{C12Y}$  is a  $(L_1 + L_2 + 1)$  square matrix of rank  $(L_2 + 1)$ . This observation allows to detect the length of the channels  $(\mathbf{h}_1, \mathbf{h}_2)$  from an estimate  $\widehat{\mathbf{C12Y}}$  of  $\mathbf{C12Y}$ .

#### 3.2. Proposed algorithms

We propose several algorithms: (i) a Root-Finding method (RF), (ii) a Sub-Space Intersection method (SSI), (iii) a method that forces the row span to have certain triangular properties (UTB), and (iv) an iterative Multidimensional Search method (MS).

(i) One can give several values to  $Z_2$  in (5), and get several functions of  $Z_1$ :  $F_{z_2}(Z_1)$ . These functions  $F_{z_2}(Z_1)$  share the roots of  $H_1(Z_1)$ :  $r_i$ , which are detected by clustering. The channel  $h_1(n)/h_1(0)$  is the inverse Z-transform of  $\prod_{i=1}^{L_1} (Z - r_i)$ , and one can build  $\mathbf{A}$ . Denoting  $\mathbf{A}^-$  the Moore-Penrose pseudo-inverse of  $\mathbf{A}$ ,  $\hat{h}_2$  is recovered via the "deconvolution":

$$\mathbf{B} = \mathbf{A}^T \cdot \mathbf{C12Y} \cdot \mathbf{A}^-. \quad (7)$$

(ii) Alternatively, one can factorize the matrix  $\mathbf{C12Y}$  in order to recover the vector  $\mathbf{h}_1$  in a similar fashion as in [8]. In the noiseless case, given that  $\mathbf{B}$  has no null eigenvalue, the matrix model (6) implies clearly that:

$$\begin{cases} \text{row}(\mathbf{A}) = \text{row}(\mathbf{C12Y}) \\ \text{col}(\mathbf{A}^T) = \text{col}(\mathbf{C12Y}) \end{cases} \quad (8)$$

Considering the singular value decomposition (SVD) of the symmetric complex matrix  $\mathbf{C12Y} = \mathbf{V}^T \cdot \mathbf{S} \cdot \mathbf{V}$ , we

define  $\hat{\mathbf{V}}$  as the  $L_2 + 1$  first rows of  $\mathbf{V}$ , associated with the  $L_2 + 1$  dominant singular values. Let  $\hat{\mathbf{V}}^{(i)}$  be the  $L_2 + 1 \times L_1 + 1$  submatrix extracted from  $\hat{\mathbf{V}}$  that gathers the columns  $i$  to  $L_1 + i$ . Then the conditions (8) are restated as:  $\mathbf{h}_1 \in \hat{\mathbf{V}}^{(i)}, \forall i \in [1, \dots, L_2 + 1]$ . Thus  $\mathbf{h}_1^H$  can be obtained by computing the dominant right singular vector of the matrix  $\mathcal{V}$  containing all  $\hat{\mathbf{V}}^{(i)}$  stacked one above the other:

$$\mathcal{V} = \begin{bmatrix} \hat{\mathbf{V}}^{(1)} \\ \vdots \\ \hat{\mathbf{V}}^{(L_2+1)} \end{bmatrix}.$$

Then the matrix  $\mathbf{B}$  can be estimated afterwards by the "deconvolution" procedure (7).

(iii) Another technique consists of forcing the UTB structure of  $\mathbf{A}$  beforehand by combining the rows of matrix  $\hat{\mathbf{V}}$ ; this is possible because of Lemma 1. Then, one extracts the  $L_2 + 1$  dimensional row vectors  $\mathbf{v}^{(i)}$  contained in the UTB matrix  $\mathcal{V}$ , and stacks them in a matrix  $\mathcal{V}$ . The rest of the procedure is identical to the previous approach (ii).

(iv) Lastly, one can perform an iterative search in the  $(L_1 + L_2(L_2 + 1)/2)$  dimensional space of the matrix product of (6) in order to find the parameters  $\theta(h_1, h_2)$  that minimize the error in the sense of the Frobenius norm:

$$\theta(h_1, h_2) = \arg \min_{\theta} \|\mathbf{C12Y} - [\mathbf{A}^T \cdot \mathbf{B} \cdot \mathbf{A}](\theta)\|_F^2$$

### 3.3. Uniqueness

**Lemma 1** *Let  $N$  and  $P$  be two positive integers. Under certain regularity conditions, any  $N \times (N + P)$  rectangular matrix  $\mathbf{M}$  can be put in UTB form by pre-multiplication by a square invertible matrix  $\mathbf{T}$ . The matrix  $\mathbf{T}$  is unique up to an invertible diagonal multiplicative matrix.*

*Proof:* The constructive algorithm is very similar to Gaussian elimination. Assume there are two matrices  $\mathbf{T}_1$  and  $\mathbf{T}_2$  such that  $\mathbf{M} = \mathbf{T}_1 \mathbf{U}_1$  and  $\mathbf{M} = \mathbf{T}_2 \mathbf{U}_2$ , where  $\mathbf{U}_1$  and  $\mathbf{U}_2$  are UTB. Then, considering the  $N$  first columns of both sides shows that the matrix  $\mathbf{T}_1 \mathbf{T}_2^{-1}$  relates two Lower Triangular (LT) matrices, and is thus LT itself. Similarly, considering the  $N$  last columns shows that  $\mathbf{T}_1 \mathbf{T}_2^{-1}$  is Upper Triangular (UT). Thus, it is diagonal, which eventually shows that  $\mathbf{T}_1$  and  $\mathbf{T}_2$  are related by a diagonal multiplicative matrix.  $\square$

**Lemma 2** *Any symmetric complex matrix  $\mathbf{C}$  can be factorized as  $\mathbf{C} = \mathbf{L} \mathbf{L}^T$ , where  $\mathbf{L}$  is lower triangular. Matrix  $\mathbf{L}$  is unique up to the post-multiplication of a diagonal matrix  $\Delta$  formed of signs  $\{\pm 1\}$ .*

**Proposition 3** *If  $\mathbf{B}$  is square full rank, and  $\mathbf{A}$  is UTB, then the decomposition of a complex symmetric matrix  $\mathbf{C} = \mathbf{A}^T \mathbf{B} \mathbf{A}$  is unique up to a multiplicative diagonal matrix.*

*Proof:* The proposition is a direct consequence of lemmas 1 and 2. It is easily seen that if  $(\mathbf{A}, \mathbf{B})$  is solution, then so is  $(\Delta \mathbf{A}, \Delta^{-1} \mathbf{B} \Delta^{-1})$ , where  $\Delta$  is any diagonal regular matrix.  $\square$

**Corollary 4** *Let  $\mathbf{B}$  be full rank symmetric complex and  $\mathbf{A}$  Töplitz UTB. When the decomposition of a symmetric matrix as  $\mathbf{C12Y} = \mathbf{A}^T \cdot \mathbf{B} \cdot \mathbf{A}$  exists, then it is unique up to a scalar multiplicative factor.*

*Proof:* From proposition 3, if  $\mathbf{A}$  is solution, then so is  $\Delta \mathbf{A}$ , with  $\Delta$  diagonal. But because  $\mathbf{A}$  is Töplitz,  $\Delta \mathbf{A}$  can be Töplitz only if  $\Delta$  is proportional to the Identity matrix.  $\square$

## 4. SIMULATIONS

In order to illustrate the Root-Finding (RF) method step by step, we first present a typical example with only RF and MS methods. Later we show a more exhaustive study with all the methods. Because we are mainly interested in direct methods, the MS is given only as a reference.

In all simulations, the input  $\mathbf{x}$  was 4-PSK. We used the real channel given by [9] ( $h_1 = [1, 0.5, -0.8, 1.6, 0.4]$  and  $h_2 = [1, 0.6, 0.6, -0.3]$ ).

**Typical example:** The input is QPSK; the number of samples is 16284 points; and the SNR is 10 dB. Figure 1.a. illustrates the clustering method. It shows all the roots calculated for different  $Z_2$ , the true roots, and the ones estimated by the method, the estimated roots (stars) are fairly accurate and match the real ones (square). Figure 1.b. shows the spectra of the real and estimated linear channels. Both estimated spectra are fairly accurate.

**Computer comparisons:** A first study showed that the estimation noise of  $\hat{C}_{12y}$  is rapidly predominant over the additive noise contribution. As expected, the Gaussian noise does not interact in the third-order moment as soon as the length of integration is long enough. So we mainly tried to estimate the influence of the number of samples. For each number of samples we took 1000 independent realizations, and the SNR is 10 dB. For each realization, we estimated  $\hat{C}_{12y}$ , on which we applied all the algorithms. Since in our case  $\mathbf{C12Y}$  is  $6 \times 6$ , the most computational intensive step is its estimation for the direct methods. Due to its iterative nature, up to several thousand of samples, the most intensive step for the MS method is the multi-dimensional search.

Figure 11 presents the influence of the integration length on the mean and variance of both estimates.

Figures II.a. and II.b. show that all methods converge to the true channel, the bias behaves well from 4096 points. The RF is the slowest method to converge to the expected value, while the MS is the fastest to converge. The SSI and the UTB follow similar patterns.

Figures II.c. and II.d. present the variances of both methods. The variances follow approximately a linear slope. It is difficult to decide which method behaves the best. One can notice that the MS has stationary performance after 64000 samples, this is because this method was implemented in a too rustic way, and it happens that a few times the MS algorithm is stuck in local minima, thus degrading the quality of the standard deviation. While not visible on the figure, the best method varies for each element of  $h_1$ , and generally around 4096 samples the best method changes. Nevertheless, above 4096 samples clearly the best method is the SSI.

The variance shows well the usual problem with High Order Statistics: in order to have consistent high-order moment estimate, the integration length must be long enough: a minimum of 8192 seems to be required here.

## 5. CONCLUDING REMARKS

Several methods have been proposed to blindly identify a linear-quadratic channel for communication applications. The idea is to use the specificities of the distribution of the inputs. The methods have shown to converge with a good accuracy, with a rather large number of samples.

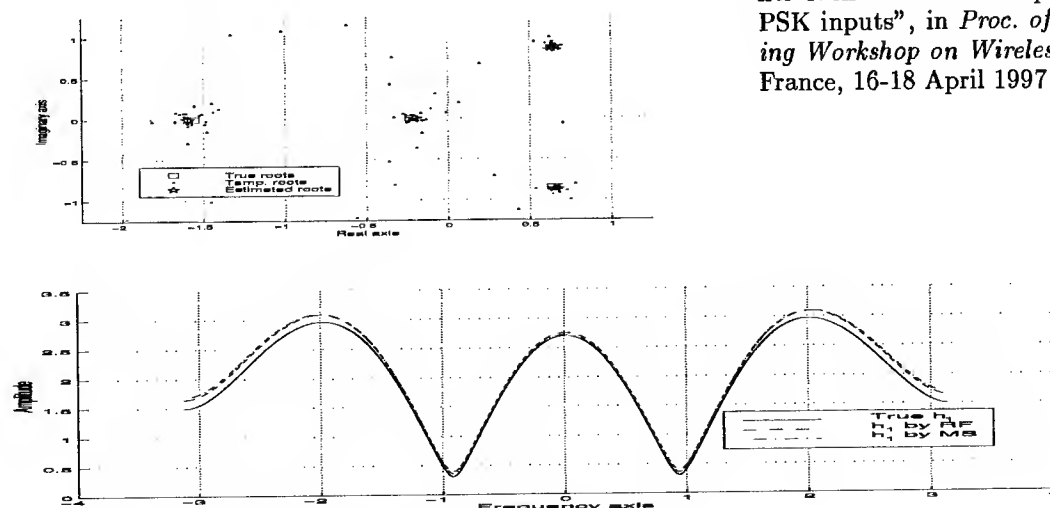


Figure I: Example of Identification: 10dB, 16284 points.

## REFERENCES

- [1] S. BENEDETTO, E. BIGLIERI, V. CASTELLANI, *Digital Transmission Theory*, Prentice-Hall Inc., New Jersey, 1987.
- [2] P. BONDON, M. KROB, "Blind identifiability of quadratic stochastic system", *IEEE trans. on Information Theory*, vol. 41, no. 1, pp. 245-254, Jan. 1998.
- [3] N. PETROCHIOS, "Elements for blind identification of non-linear channels", supervision by G. Giannakis, Master's thesis, ENSEA / ENS Lyon, Sept. 1996, In archive of ENSEA, France.
- [4] E. J. POWERS, S. IM et al., "Applications of hos to nonlinear hydrodynamics", in *IEEE-ATHOS Workshop on Higher-Order Statistics*, Begur, Spain, 12-14 June 1995, pp. 414-418.
- [5] W. J. RUGH, *Nonlinear System Theory*, Johns Hopkins Univ. Press, Baltimore, MD, 1981.
- [6] M. SCHETZEN, *The Volterra and Wiener Theories of Nonlinear Systems*, Wiley, New York, 1980.
- [7] H-Z. TAN, Z-Y. MAO, "Blind identifiability of quadratic non-linear systems in higher-order statistics domain", *Int. Jour. Adapt. Control Signal Processing*, vol. 12, pp. 567-577, 1998.
- [8] A. J. van der VEEN, S. TALWAR, A. PAULRAJ, "A subspace approach to blind space-time signal processing for wireless communication systems", *IEEE trans. on Signal Processing*, vol. 45, no. 1, pp. 173-190, Jan. 1997.
- [9] G. T. ZHOU, G. B. GIANNAKIS, "Nonlinear channel identification and performance analysis with PSK inputs", in *Proc. of 1st IEEE Signal Processing Workshop on Wireless Communications*, Paris, France, 16-18 April 1997, pp. 337-340.

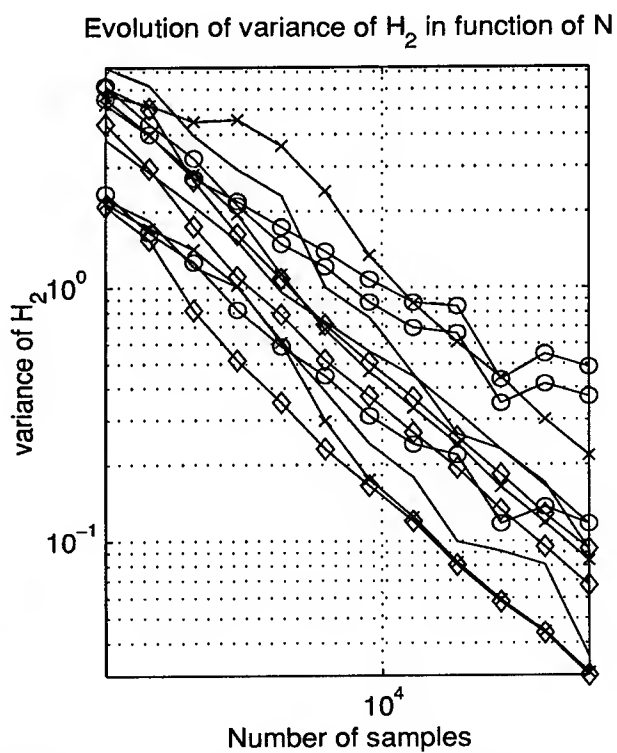
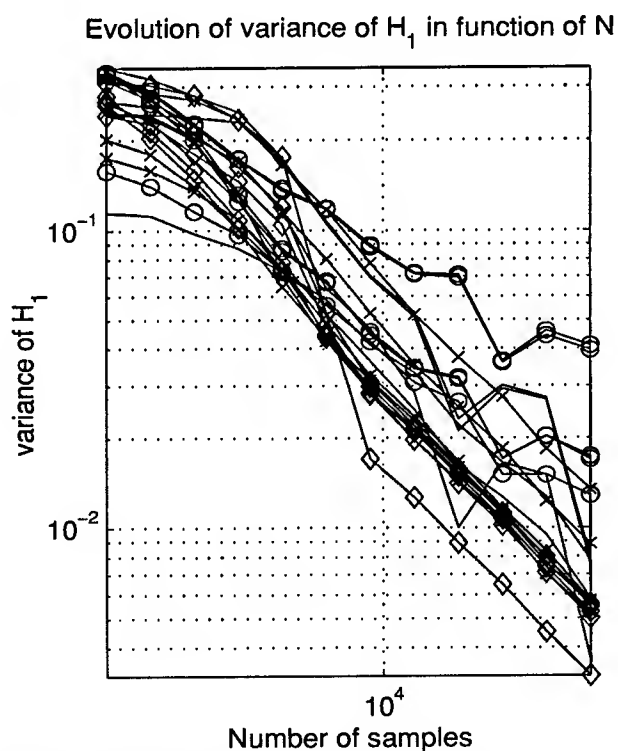
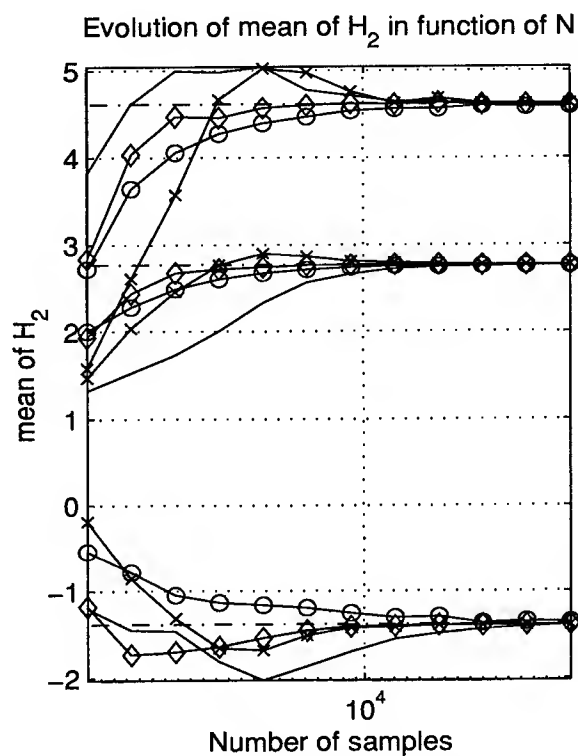
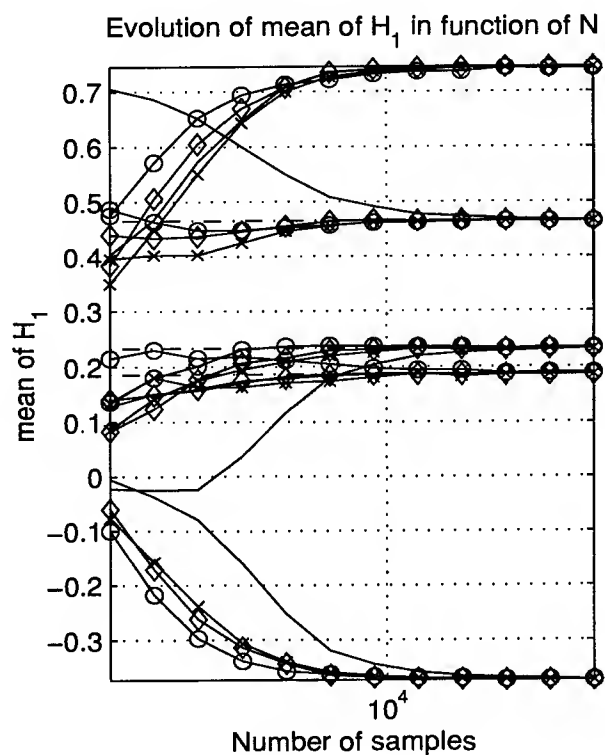


Figure II: Means and standard deviations for all methods with 1000 independent realizations. Simple line: RF, x-: UTB, d-: SSI, o-: MS.

# JOINT CHANNEL ESTIMATION AND DETECTION FOR INTERFERENCE CANCELLATION IN MULTI-CHANNEL SYSTEMS

*Cristoff Martin and Björn Ottersten*

The Department of Signals, Sensors & Systems  
Royal Institute of Technology (KTH)  
SE-100 44 Stockholm, Sweden

## ABSTRACT

Interference from other users and interference due to multipath propagation limit the capacity of wireless communication networks. As the number of users and the demand for new services in the networks increases, co-channel interference will be a limiting factor.

This paper proposes an iterative structured multi-channel receiver algorithm that jointly estimates the communication channels and desired data while canceling interference. A general way of adding training redundancy to a data frame is also introduced.

From simulations the proposed method is shown to achieve low bit error rates also in the presence of strong interference. These simulations also show that by distributing the training information in a data burst elaborately, further improvements in performance are achievable.

## 1. INTRODUCTION

During the last decades, a rapid development in mobile communications has occurred. The seemingly ever increasing number of users and services has caused equally increasing demand for capacity and reliability. Because of the physical limitations of radio communications and the limited bandwidth available these demands are difficult to meet.

One of the factors that limits capacity is the interference from other users, *Co-Channel Interference* or CCI. The problem is further complicated by the fact that in realistic wireless communication systems there will always be some amount of multi-path propagation causing *Inter-Symbol Interference* or ISI. Thus, by developing receivers that can handle these kinds of interference, the capacity and reliability in the wireless network can be increased. One way of combating interference is through the use of antenna arrays, thus creating a multi-channel system. The receiver systems considered in this paper are all multi-channel.

This paper considers an iterative algorithm that at the same time it is rejecting interference also estimates transmitted data and baseband transmission channels. The proposed receiver is semi-blind, i.e., it uses training information available for the desired user.

Several other approaches have been taken to reject interference. Iterative Least Squares with Projection (ILSP) is introduced in [1, 2]. ILSP is a blind method to separate several co-channel signals using the *Finite Alphabet* (FA)

property of digital communication signals. However it does not handle ISI nor does it handle training information in a natural fashion. The method presented herein is similar to ILSP but taking ISI and training information into account as well. In [3] an interference rejection algorithm is presented that by using ILSP, oversampling and an extra processing step is able to also handle ISI. Another method similar to ILSP is proposed in [4], this method also handles training sequences and ISI. However it does not handle the structure imposed by the ISI. Another class of interference rejection algorithms are subspace methods. These use algebraic subspace properties to reject interference based on second order statistics. An example of such a method used for comparison in this paper can be found in [5].

## 2. DATA MODEL

An  $L$  element antenna, with symbol spaced base band sampling is considered. For simplicity only one desired user and one interfering user is considered (even though the data model and proposed receiver algorithm easily can be extended to multiple users and interferers). The interferer is assumed to be using the same modulation scheme as, and be burst synchronized with, the desired user. Within a burst the user and the interferers send one data frame consisting of  $N$  symbols of which  $N_D$  symbols are unknown data and the rest are used for training purposes. The radio channels between the transmitters and the receiving antennas are assumed to be time invariant within one data frame. It is also assumed that the transmission process between the transmitter and the receiver, including the effects of the transmitter and receiver filters can be modeled as a FIR filter of length  $M$ . It is then possible to model the received data as

$$\mathbf{X} = \mathbf{H}\mathbf{S} + \mathbf{G}\mathbf{D} + \mathbf{V}. \quad (1)$$

Where  $\mathbf{X}$  (which is  $L \times (N + M - 1)$ ) contains the data received by the antenna array. The channel matrices,  $\mathbf{H}$  and  $\mathbf{G}$  (both  $L \times M$ ), describe the transmission process between the desired user and the interferer respectively. The transmitted data is contained in  $\mathbf{S}$  and  $\mathbf{D}$  ( $M \times (N + M - 1)$ ) while  $\mathbf{V}$  models additive noise. The received data matrix  $\mathbf{X}$  is organized as  $\mathbf{X} = [\mathbf{x}(1) \ \mathbf{x}(2) \ \dots \ \mathbf{x}(N + M - 1)]$  where  $\mathbf{x}(n)$  is a column vector containing the the data output from the array at the  $n$ th sampling instant. To exemplify the organization of the data matrices, the data matrix

of the desired user is

$$\mathbf{S} = \begin{bmatrix} \boxed{\mathbf{s}^T} & 0 & \dots & 0 \\ 0 & \boxed{\mathbf{s}^T} & \dots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & \boxed{\mathbf{s}^T} \end{bmatrix}. \quad (2)$$

Where  $\mathbf{s}$  is a vector containing the data symbols transmitted in one frame. From (2) the structure of the data matrices becomes obvious. In order to achieve good performance a receiver algorithm must preserve this structure.

### 3. PROBLEM FORMULATION

The problem of estimating the unknown data vectors and channel matrices is considered. It is assumed that training information is available for the desired user while it is unknown for the interferer. The transmission of the data is disturbed by spatially and temporally additive white complex Gaussian noise.

The goal is to find the maximum likelihood estimates of  $\mathbf{H}$ ,  $\mathbf{S}$ ,  $\mathbf{G}$  and  $\mathbf{D}$ . That is, the  $\mathbf{H}$ ,  $\mathbf{S}$ ,  $\mathbf{G}$  and  $\mathbf{D}$  that minimizes

$$\|\mathbf{X} - \mathbf{H}\mathbf{S} - \mathbf{G}\mathbf{D}\|_F^2 \quad (3)$$

taking the finite alphabet property of the signals into account. Note that given the data symbols, the criterion is quadratic in the channel matrices. After rewriting this norm as

$$\|\mathbf{X} - \mathbf{H}\mathbf{S} - \mathbf{G}\mathbf{D}\|_F^2 = \left\| \mathbf{X} - \begin{bmatrix} \mathbf{H} & \mathbf{G} \end{bmatrix} \begin{bmatrix} \mathbf{S} \\ \mathbf{D} \end{bmatrix} \right\|_F^2 \quad (4)$$

it can be minimized with respect to  $\begin{bmatrix} \mathbf{H} & \mathbf{G} \end{bmatrix}$ ,

$$\begin{aligned} \widehat{\begin{bmatrix} \mathbf{H} & \mathbf{G} \end{bmatrix}} &= \mathbf{X} \begin{bmatrix} \mathbf{S} \\ \mathbf{D} \end{bmatrix}^\dagger \\ &= \mathbf{X} \begin{bmatrix} \mathbf{S} \\ \mathbf{D} \end{bmatrix}^* \left( \begin{bmatrix} \mathbf{S} \\ \mathbf{D} \end{bmatrix} \begin{bmatrix} \mathbf{S} \\ \mathbf{D} \end{bmatrix}^* \right)^{-1}. \end{aligned} \quad (5)$$

Where  $\mathbf{A}^\dagger$  denotes the pseudo inverse of  $\mathbf{A}$ . After having resubstituted  $\widehat{\mathbf{H}}$  and  $\widehat{\mathbf{G}}$  into (4) a minimization criterion only depending on  $\mathbf{S}$  and  $\mathbf{D}$  is achieved,

$$\min_{\mathbf{S}, \mathbf{D}} \left\| \mathbf{X} \mathbf{P}_{\begin{bmatrix} \mathbf{S} \\ \mathbf{D} \end{bmatrix}}^\perp \right\|_F^2 \quad (6)$$

Where  $\mathbf{P}_{\mathbf{A}}^\perp = \mathbf{I} - \mathbf{A}^* (\mathbf{A} \mathbf{A}^*)^{-1} \mathbf{A}$  and  $\mathbf{I}$  is the identity matrix. It is now possible to find the global minimum by enumerating over all possible  $\mathbf{S}$  and  $\mathbf{D}$  using their FA-property and known training information while maintaining the structure of the matrices. The enumerating however is of exponential complexity which makes this enumerating impossible also for modest data frame sizes. The following sections consider a suboptimal method that attempts to minimize (3) with less computational complexity.

### 4. PROPOSED ALGORITHM – OUTLINE

The algorithm proposed in this paper takes an iterative approach to minimize (3) while maintaining the structure of the data matrices (see (2)). Known training information is also taken into account. The iterative procedure of the proposed algorithm is similar to the ILSP algorithm proposed in [1, 2].

Assuming that initial estimates of the data sequences are available the method can be outlined as

1. Assume that the estimated data sequences are correct. The norm (3) is now quadratic in  $\mathbf{H}$  and  $\mathbf{G}$  and it is easy to estimate the channel matrices.
2. Rewrite the norm (3) so that it can be minimized in a way that maintains the structure of  $\mathbf{S}$  and  $\mathbf{D}$  and takes available training information into account.
3. Now, assume that the estimated channel matrices are correct. The norm (3) becomes quadratic in  $\mathbf{S}$  and  $\mathbf{D}$  if we relax the FA-property. Thus, it is possible to estimate the unknown data symbols by solving a linear set of equations.
4. Project the data on its finite alphabet.
5. Repeat the steps above until convergence.

If the initial data estimates are good enough the method will in general converge to the desired global minimum of (3) and the initial data estimate are improved.

The proposed method also makes it possible to generalize how the training information is added to the data sequence. This is considered in the following section. A more detailed description of the algorithm can be found in section 6.

### 5. GENERALIZED TRAINING USING CODE MATRICES

When a training sequence is added to a data frame it is usually either simply inserted in the beginning or at the middle of the data frame. Here a more general way of adding the training data is introduced by the affine mapping

$$\mathbf{s} = \mathbf{C}_1 \mathbf{s}_d + \mathbf{C}_0. \quad (7)$$

Where  $\mathbf{s}$  ( $N \times 1$ ) contains the data to be transmitted (data and training information),  $\mathbf{s}_d$  ( $N_d \times 1$ ) contains the data without training information.  $\mathbf{C}_1$  ( $N \times N_d$ ) and  $\mathbf{C}_0$  ( $N \times 1$ ) are *Code Matrices* that add training information (and possibly error correcting redundancy) to the data.

It is obvious that the code matrices can be chosen so that training information is added to the data sequence in the conventional way described above. However this also provides the opportunity of adding training information more elaborately. For example the training information can be distributed over the entire data sequence.

### 6. PROPOSED ALGORITHM – DETAILS

The steps of the proposed algorithm outlined in section 4 are presented in more detail in this section. It is assumed that an initial estimate of the unknown user data and the

interferer data is present. Further it is assumed that the code matrices  $C_0$  and  $C_1$  are known for the desired while they are not available for the interferer.

### 6.1. Estimating the Channel Matrices

If we assume the estimated data sequences to be correct a least squares estimate of the channel matrices can be found as (see (5))

$$\begin{bmatrix} \widehat{H} & \widehat{G} \end{bmatrix} = X \begin{bmatrix} S \\ D \end{bmatrix}^\dagger.$$

### 6.2. Maintaining the Structure of the Data Matrices

To maintain the structure of the data matrices while estimating them the norm (3) must be rewritten. This can be achieved using properties of the vec operator and the Kronecker product. Letting  $\text{vec}$  denote the vec operator,  $\otimes$  denote the Kronecker product and  $I$  denote the identity matrix, this rewriting can be done in a few steps as follows,

$$\begin{aligned} \text{vec}\{X - HS - GD\} &= \text{vec} X - (I \otimes H) \text{vec} S \\ &\quad - (I \otimes G) \text{vec} D. \end{aligned} \quad (8)$$

To simplify notation, let  $\Phi_H = I \otimes H$ ,  $\Phi_G = I \otimes G$ , and  $x = \text{vec} X$ . Also, the  $(NM \times N)$  selection matrix  $\Psi$  is defined. The matrix  $\Psi$  consists of zeros and ones and takes a data vector to a vectorized data matrix, i.e.  $\text{vec} S = \Psi s$  and  $\text{vec} D = \Psi d$ . Now, (8) can be rewritten as

$$\begin{aligned} \text{vec}\{X - HS - GD\} &= x - \Phi_H \Psi s - \Phi_G \Psi d \\ &= x - \Phi_H \Psi C_1 s_d \\ &\quad - \Phi_H \Psi C_0 - \Phi_G \Psi d \\ &= x - \Phi_H \Psi C_0 \\ &\quad - [\Phi_H \Psi C_1 \quad \Phi_G \Psi] \begin{bmatrix} s_d \\ d \end{bmatrix}. \end{aligned} \quad (9)$$

where the middle step follows from (7). By using (9) the norm (3) can now be minimized with respect to the data, while maintaining the structure of the data matrices  $S, D$ .

### 6.3. Estimating the Received Data

By using (9) and assuming the estimated channels to be correct we now obtain continuous estimates of the unknown data vectors  $s_d$  and  $d$ . This can be done much in the same way as the estimation of the channel matrices which results in

$$\begin{bmatrix} \widehat{s}_d \\ \widehat{d} \end{bmatrix} = [\Phi_H \Psi C_1 \quad \Phi_G \Psi]^\dagger (x - \Phi_H \Psi C_0). \quad (10)$$

The unknown data can be estimated by projecting the continuous data estimates to the finite alphabet in use.

Finally the three steps above are iterated until convergence is reached. If the initial estimates are good enough they are in general improved.

## 7. PRELIMINARY RESULTS

To give some insight to the kind of performance that the proposed algorithm might offer, simulations have been conducted and the results from these are presented in this section. In order to offer some comparison with previous work, the structured subspace receiver described in [5] was simulated under the same conditions and results from these simulations are provided.

Two different sets of code matrices were used (see section 5). One conventional with all the training symbols in the beginning of the sequence and one with the training symbols spread over the entire sequence. In the simulations of the structured subspace receiver the entire training sequence was located in the beginning of the data frame.

In all cases an  $L = 4$  antenna system was considered. An antipodal binary modulation scheme was employed (this would for example correspond to BPSK).

To model the transmission process (the transmitter/receiver filters and the radio channel) a two tap FIR channel model was used. The channels were assumed independent from antenna to antenna and to simulate Rayleigh fading the channel taps were independently drawn from a complex Gaussian distribution.

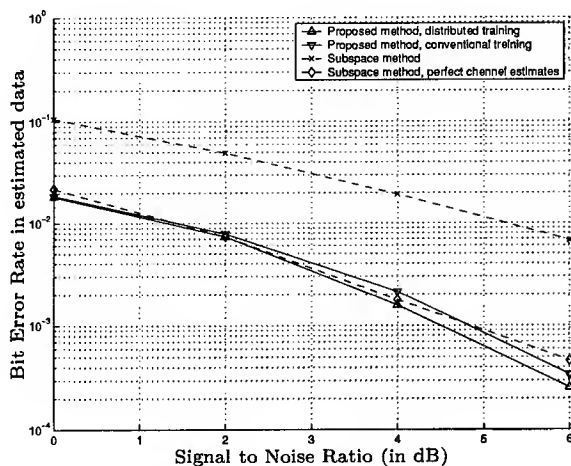
In the simulations it was assumed that the length of the channel impulse responses,  $M$ , and the number of transmitters,  $U$ , are known or have been correctly estimated.

To offer some idea about what the achievable performance would be, a simple initialization scheme was employed. Interferer data was initialized with its continuous solution (of the minimization of the norm (4), ignoring the structure of the data matrices, see e.g [2]) projected to the finite alphabet in use. The desired user data was initialized with random data symbols. Received sequences where the resulting norm (3) was smaller than the true norm (the norm (3) achieved using the true data and channel matrices) plus one standard deviation of the norm were kept while received sequences not fulfilling this criteria were identified as outliers.

In figure 1 the bit error rate performance of the proposed method as a function of the *Signal to Noise Ratio* (SNR) is shown. The desired user is disturbed by a single interferer. The *Signal to Interference Ratio* (SIR) in these simulations was -10 dB. The results from the simulated proposed method are compared with the structured subspace method with estimated channels and with known channels. Also, the two different sets of training matrices (described above) are compared. The data frames consist of 57 symbols of which 42 are data symbols and the rest are used for training purposes. At these conditions the proposed method performs well on par with the structured subspace method using perfect channel estimates. The structured subspace method by itself needs longer training sequences in order to perform well (see figure 4). The distributed training information offers slightly better performance than the conventional training sequence. Even though the difference in performance is small this is interesting as both these data distributions use the same number of training and data bits. Only how they are distributed differ.

To explore the loss in performance due to the interference, the proposed method was simulated with and with-





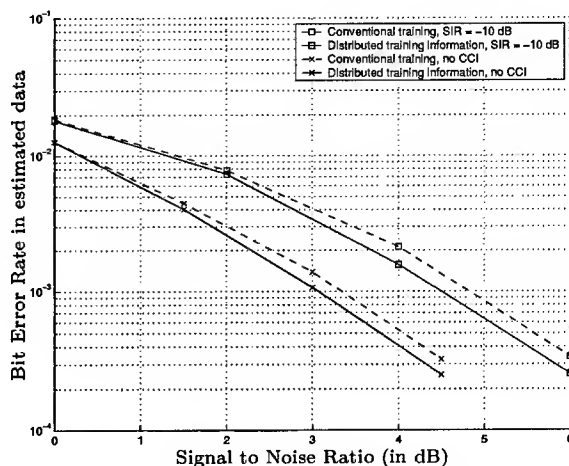
**Figure 1:** Performance with a single -10 dB interferer present.

out an interferer. Other than that the simulated conditions were identical to the previous simulation. The results from these simulations are shown in figure 2. As can be seen from the graph, at an SNR of 4 dB the loss is approximately 1.5 dB, both with the conventional training sequence and with the distributed training information. Again slightly lower bit error rates were achieved when the distributed training information was used compared to the more conventional training data distribution.

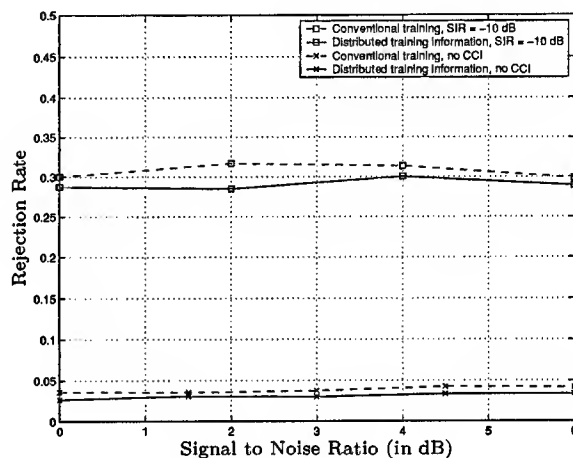
The number of data frames not converging to a norm small enough, the *rejection rate*, was also measured under the same conditions as in the previous simulations. Figure 3 shows the results from these measurements. As can be seen from the graph, when there is CCI present the rejection rate becomes quite high and it would be desirable to use a better initialization method.

The effects of the length of the training sequence was also given some attention. Once again the proposed algorithm with the two different training distributions and the structured subspace method (found in [5]) were compared. Figure 4 shows the bit error rate of the desired data sequence as a function of the number of training symbols and figure 5 shows the rejection rate as a function of the number of training symbols. These simulations were performed at an SNR of 4 dB, with and without a single -10 dB co-channel interferer. The number of data symbols in each frame remained 42. From figure 4 it can also be seen that the proposed method is less sensitive to short training sequences than the method used for comparison. Figure 5 shows that the number of rejected sequences increases fast when the number of training symbols drops below 15. It seems likely that the convergence criteria might affect simulated bit error rates when the number of training symbols becomes smaller than that.

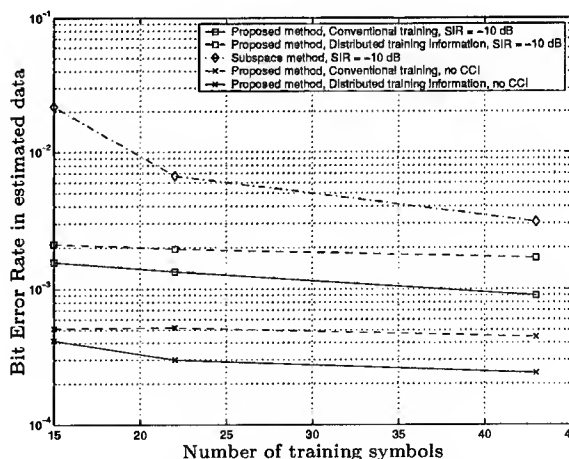
As can be seen from the results above the proposed method is showing promising performance. However there are still several issues that require further investigation. For example, in its current implementation the proposed receiver algorithm is computationally expensive. Also the ro-



**Figure 2:** Performance lost due to interference.



**Figure 3:** Rejection rates as functions of the SNR.



**Figure 4:** Error rates at an SNR of 4 dB.

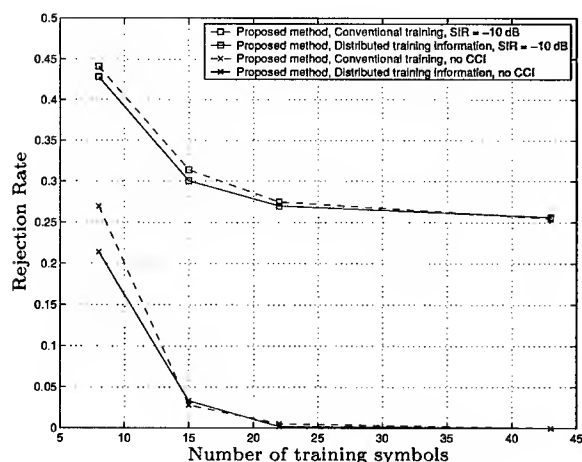


Figure 5: Rejection rates at an SNR of 4 dB.

business to model errors and initialization are other issues that deserve more attention. More general forms of training information where the data is confined to more general affine mappings can easily be considered with the proposed method.

## 8. CONCLUSIONS

Herein, we have presented a interference cancellation method that can be applied to multi-channel data. Training information from the desired user is exploited and the communication channels are jointly estimated together with the unknown data symbols of both the desired user and the interference. This method can easily treat general forms of training information and a simple example with distributed training information was shown to give improved performance compared to a block of training data.

## 9. REFERENCES

- [1] S. Talwar, M. Viberg, and A. Paulraj, "Blind estimation of multiple co-channel digital signals using an antenna array," *IEEE Signal Processing Letters*, vol. 1, February 1994.
- [2] S. Talwar, M. Viberg, and A. Paulraj, "Blind separation of synchronous co-channel digital signals using an antenna array – part I: Algorithms," *IEEE Transactions on Signal Processing*, vol. 44, pp. 1184–1197, May 1996.
- [3] A.-J. van der Veen, S. Talwar, and A. Paulraj, "Blind identification of FIR channels carrying multiple finite alphabet signals," in *Proc. of ICASSP*, vol. 2, pp. 1213–1216, 1995.
- [4] J. Laurila, R. Tschöfen, and E. Bonek, "Semi-blind space-time estimation of co-channel signals using least squares projections," in *Proceedings of the Vehicular Technology Conference, 1999. VTC 1999 - Fall.*, vol. 3, pp. 1310 – 1315, Sept 1999.

- [5] G. Klang and B. Ottersten, "Channel estimation and interference rejection for multichannel systems," in *Proceedings of the 32th Asilomar Conference on Signals, Systems and Computers*, (Pacific Grove, CA, USA), nov 1998.

# A SPATIAL CLUSTERING SCHEME FOR DOWNLINK BEAMFORMING IN SDMA MOBILE RADIO

Wen-Jye Huang and John F. Doherty

Department of Electrical Engineering  
The Pennsylvania State University  
University Park, PA 16802  
E-mail: {wxh148,jfdoherty}@psu.edu

## ABSTRACT

In this paper we proposed a new approach that clusters mobile users before downlink beamforming and broadens beams and nulls within the beamforming calculation. We first investigate the broadening beamforming scheme to alleviate inaccuracies in DOA estimation. Next we exam how to group the mobile users, with the constraint of separation angle, to enhance downlink beamforming. Simulations show that the downlink beamforming complexity is decreased dramatically with limited performance loss.

## 1. INTRODUCTION

Owing to the rapid growth demand in the mobile communication, the current capacity of mobile communication faces a severe challenge during peak usage. To remedy the capacity limitation, research on space-division multiple-access (SDMA), which increases system capacity and decreases co-channel interference, has been investigated.

A basic idea of SDMA is to spatially separate the mobile users, which allows reuse of limited radio resources, such as frequency, time, or code slot within a cell. SDMA relies on the application of an adaptive array antenna at the base station to form multiple beam patterns, which serve multiple user traffic channels. Therefore the capacity of the system can be increased.

Prior research shows that implementing SDMA on the downlink increases the channel capacity [1], [2], [3]. One simple SDMA approach uses the DOA estimated from uplink data and forms the spatial signature for downlink transmission. However, in urban environments, angular spreads (AS) could be up to  $15^\circ$  [4], which means the estimated downlink beamforming pattern may degrade system performance due to narrow, misaligned nulls. In addition, if the user DOAs are

not well separated, SDMA cannot provide much system performance improvement. Furthermore, the downlink beamforming algorithm needs extensive computation power to solve a nonlinear optimization problem involving a nonlinear constraint weight vector for every user [5]. This limits the applicability of this approach for low complexity, real-time operation.

This paper proposes a new approach that clusters (groups) mobile users before the downlink beamforming calculation. This approach alleviates the computational complexity problem and the spatial separability problem. The algorithmic block diagram is shown in Figure 1. By carefully choosing AS and forming the same beamforming weight vector  $w_{group}$  to the same group, the simulation results show that the clustering scheme is within 3 dB of the conventional method, with a dramatic decrease in computational complexity.

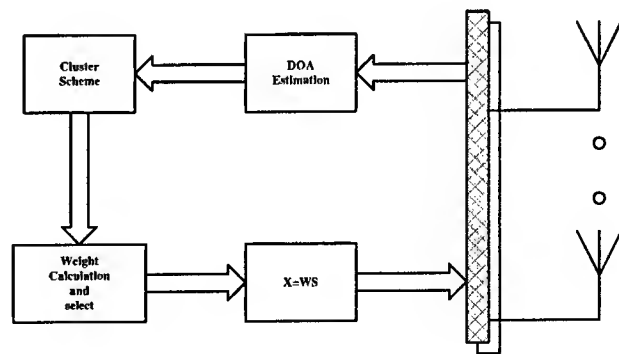


Figure 1: New Cluster Algorithm for Downlink Beamforming

## 2. DATA MODEL

We assume that  $K$  users are served within the same cell by the base station with a uniform linear array Antenna

(ULA) consisting of  $M$  identical, omnidirectional sensors, equally spaced at distance  $d$ . A narrowband signal model is assumed and the baseband signal received at time  $t$  with  $L_k$  paths for the  $k$ th user is:

$$x(t) = \sum_{k=1}^K \sum_{l=1}^{L_k} A_{kl} a(\theta_{kl}, f_u) s_k(t - \tau_{kl}) + n(t) \quad (1)$$

where  $n(t)$  is spatially and temporally white Gaussian noise and the array steering vector  $a(\theta, f_u)$  is given by

$$a(\theta, f_u) = [1, e^{-j2\pi d \frac{f_u}{c} \sin \theta}, \dots, e^{-j2\pi d \frac{f_u}{c} (M-1) \sin \theta}]^T \quad (2)$$

where  $A_{kl}$  is the amplitude of the  $l^{th}$  path of the  $k^{th}$  user,  $s_k(t)$  is the baseband signal transmitted at the  $k^{th}$  mobile and  $\tau_{kl}$  is its corresponding delay.

From the received uplink signal, it is possible to estimate the spatial covariance matrix, which contains the directional information of the mobile radio channel (dominant DOAs  $\theta_{kl}$ ) and corresponding power for each user. It can be written as following:

$$R_k = \sum_{l=1}^{L_k} A_{kl}^2 a(\theta_{kl}, f_d) a^H(\theta_{kl}, f_d) \quad (3)$$

Similarly we define the interference covariance matrix  $Q_k$  as

$$Q_k = \sum_{i \neq k} R_i + \sigma_N^2 I \quad (4)$$

where  $\sigma_N^2$  and  $I$  denote the white noise variance and  $M \times M$  identity matrix, respectively.

The goal of downlink beamforming is to design a weight vectors  $w_{kd}(f_d; t)$  to transmit the constraint power to the desired user and to minimize the transmitted energy to the undesired user. In another word we want to maximize the SINR (Signal to Noise plus Interference Ratio) for the  $k$ th user [6].

$$w_{kd} = \arg \max_{w_{kd}} \frac{w_{kd}^H R_k w_{kd}}{w_{kd}^H Q_k w_{kd}} \quad (5)$$

The solution of (5) is proportional to the generalized eigenvector of matrix pair  $[R_k, Q_k]$  [3]

$$w_{kd} = e_{dk}^{[\max]} \sqrt{\frac{\chi_k}{e_{dx}^{[\max]H} R_k e_{dx}^{[\max]}}} ; \text{ if } w_{kd}^H R_k w_{kd} = \chi_k \quad (6)$$

### 3. TARGET AND NULL BROADENING

The existence of angular spreads (AS) causes DOA estimation error, which adversely affects the downlink beamforming process. The SINR degrades because the maximum transmitted power is not directed at the desired user, or because the nulls pointed towards to the cochannel users are too narrow. One method presented in this section will make the SINR more robust to DOA estimation error. The angular spread based approach [7], [8] can steer a broad range of beam patterns towards users of interest, or nulls toward the cochannel users. A modified version of interference covariance matrix can be written as:

$$R_k^{\sim} = R_k \odot S_{\max} \quad (7)$$

$$Q_k^{\sim} = Q_k \odot S_{\max} \quad (8)$$

$$\text{with } [S_{\max}]_{pq} = e^{-2[\pi \frac{d}{\lambda_d} (p-q)]^2 \sigma_{\max}^2}$$

where  $\odot$  and  $[.]_{pq}$  denote the Schur Hadamard element-by-element matrix product and the  $pq$ th element of a matrix, respectively. The variable  $\sigma_{\max}^2$  quantifies the angular spreads (AS) of the corresponding DOAs.

By using target and null broadening technique in the downlink beamforming, the design of beamformers are more robust in the mobile communication environment. In addition, the beamforming weights are valid for a longer time with less calculations required [8]. Figure 2 shows the beam pattern with and without the broadening technique. It is clear that by applying the broadening technique, the narrow nulling interference problem is solved. Although it introduces some increase of the SINR perturbation, the worse case effect of DOA estimation error is still negligible [6].

### 4. GROUPING AND DOWNLINK BEAMFORMING ALGORITHM

Two conditions limit the performance and capacity of SDMA systems:

1. Users that share same channel allocation are co-located, within the resolution of the beam pattern;
2. Co-channel, co-located users have disparate powers, causing the so-called "near-far problem."

A proposed solution to the near far problem is grouping the mobile users within power classes before downlink beamforming [9].

Utilizing the advantage of the target and null broadening method, and the existence of angular spreading

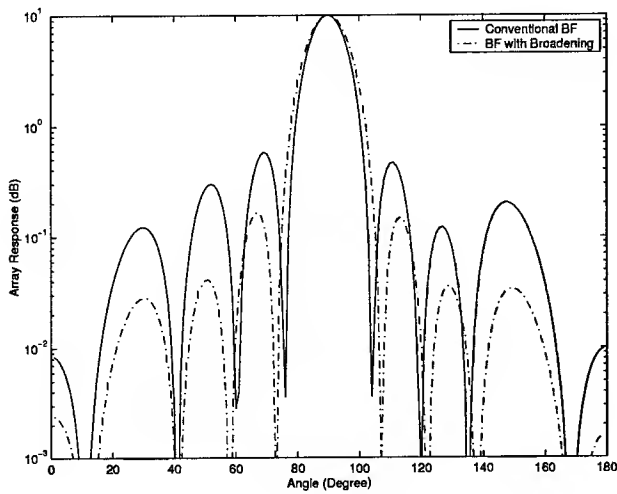


Figure 2: Conventional Beamforming vs. Beamforming with Broadening Target and Null Technique with Target at 90° and Null at 40°

(AS), we propose a grouping algorithm that is constrained to angle separation with location in a cell. By grouping all the users in a cell before downlink beamforming and selective calculation for downlink beamforming weight in a group, the computational complexity for the base station is decreased dramatically with a tolerable performance loss.

The basic approach of grouping and downlink beamforming calculation algorithm within a cell is the following:

1. Determine the angle separation  $\Delta\theta$  for each group, typically use the angular spreading (AS) as a parameter;
2. Assign users to same group if ( $\Delta\theta < AS$ );
3. Determine the representative angular for each group, typically choose the highest energy interference source within a group as a representative;
4. Calculate the downlink beamforming weight  $w_{new}$  for each group;
5. Apply the weight  $w_{new}$  for each user in the same group.

We use a simulation with  $M=8$  uniform linear antenna with half wavelength inter-element spacing to verify that the performance loss is acceptable for the above algorithm. Consider  $N=4$  sources, one signal-of-interest (SOI) and three signal-of-non-interest (SONI), with initial SOI DOA of 90° and DOA's of SONI at 40°, 120° and 140°. Figure 3 compares SINR error

for conventional beamforming and the target and null broadening technique.

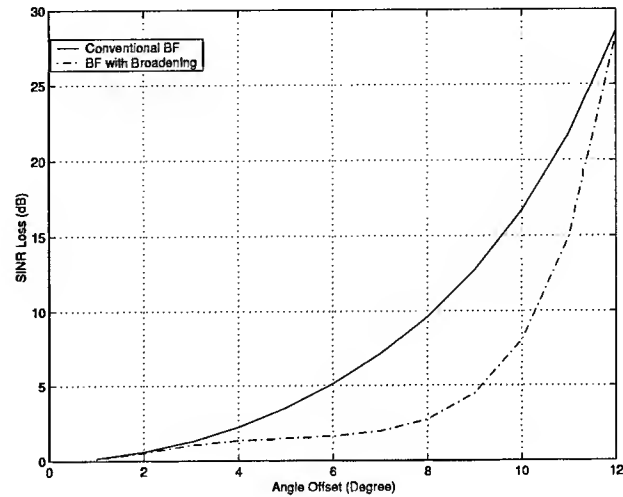


Figure 3: Downlink SINR comparison for conventional beamforming method and beamforming using the broadening technique.

From Figure 3, it is clear that if users are geometrically close enough, in this case  $AS \leq 8^\circ$ , we can reuse the same downlink weight  $w_{new}$  to save calculations in base station with an acceptable trade-off 3dB SINR loss, in this case. However, if we account for interference source spreading angles, which are due to the narrow nulls of traditional beamforming, the performance loss due to angle spreading towards the co-channel users is large. Figure 4 shows the performance loss due to offset targeting the co-channel users for the previous simulation scenario. It is obvious that the broadening technique reduces performance loss due to co-channel angle spreading.

We use a simulation to demonstrate the complexity savings of the grouping method. Figure 5 shows the performance under different angle spreading, where users are uniformly distributed by angle in a cell.

The results shown in Figure 3 and Figure 5 indicate that, with proper grouping user within a cell, it is possible to save more than 50% of downlink beamforming computational complexity with limited SINR performance loss.

## 5. SIMULATIONS

The simulations model a system that uses a linear array antenna with  $M = 8$  antennae and half wavelength inter-element spacing and  $N = 25$  mobile users uniformly distributed from  $[0 \pi)$  within a cell. Figure

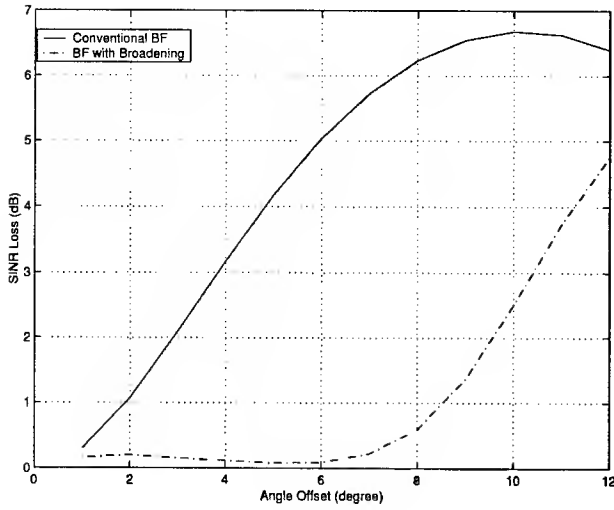


Figure 4: Performance loss due to co-channel users angle offset.

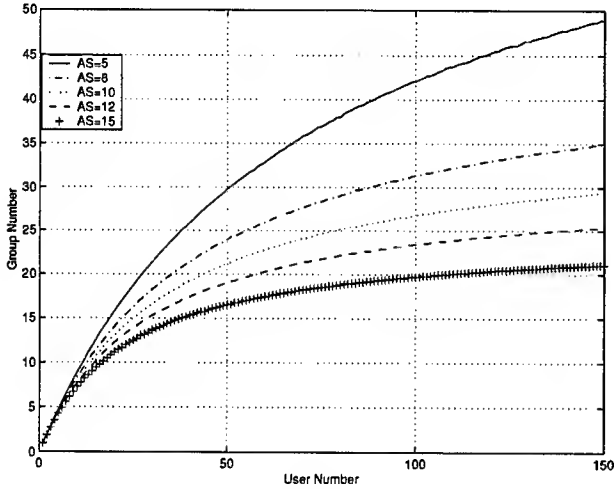


Figure 5: Group number vs. user number under various angle spreading conditions.

6 shows the block diagram for conventional downlink beamforming and the flow chart for the grouping algorithm.

Based on Figure 6, Table 1 addresses, under the simulation environment model, the computational load for each block.

It is obvious that the proposed method needs only one-third of typical base station complexity for calculation  $RQ$  and  $w_{down}$ . From the entire system viewpoint, the new method reduces the computational complexity needed in the base station for SDMA applications by

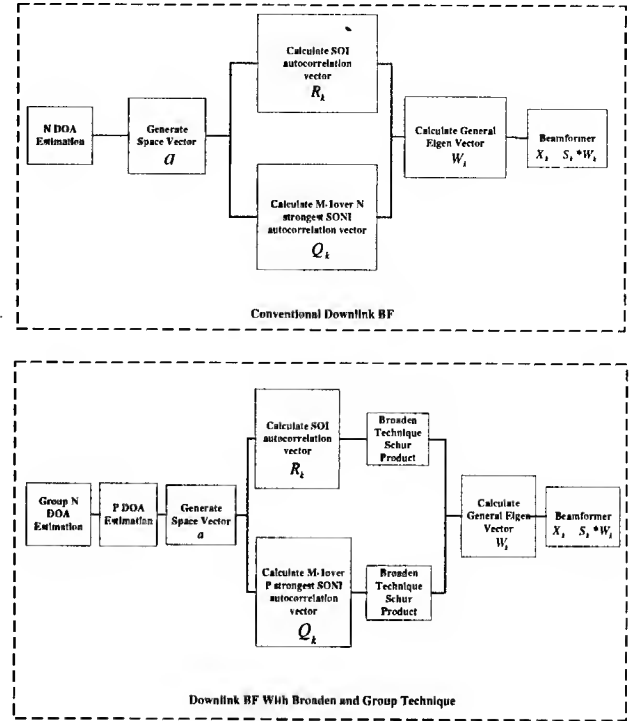


Figure 6: Block Diagram for Conventional Dowlink BF Algorithm and Algorithm with Broadening Technique

approximately 50%.

Figure 7 shows the performance of grouping plus broadening target and nulls scheme, assuming that angle spreading exists on all sources (desired user and cochannel interference). The worse scenario is target and nulls not coincident with the estimated DOAs are at maximum offset,  $AS = 8$ . Figure 7 shows that worse case SINR loss decreases substantially by using group and broadening scheme.

Combining the results of Figure 6 and Figure 7 indicates the efficacy of the new approach. By grouping mobile user in a cell, and using the broadening target and nulls technique, the downlink beamforming calculation is reduced by approximately 50%, with acceptable performance loss.

## 6. CONCLUSION

In this paper, we have studied the grouping and broadening target and nulls technique for downlink beamforming in mobile communication systems. Computer simulations show that the benefit of grouping users not only can alleviate the DOA estimation error problem, but also can offer robust beamforming performance in the present of source movement [8]. Moreover, the computation complexity in the base station is de-

	BF with broadening	Conventional BF	Calculation Effort
$R$	8	25	(3)
$Q$	8	25	(4)
$a(\theta)$	8	25	(2)
$w$	8	25	(6)
$X$	25	25	$X = S * W$
Schur Product	$8*2$	0	$\odot$
Decision	25 weight Select	0	

Table 1: Computational Effort Comparison for Conventional BF and BF with Group and Broadening Technique

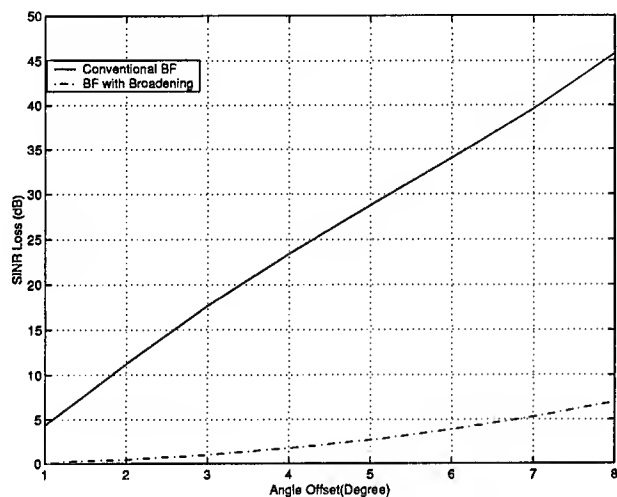


Figure 7: Simulation Result Under  $N=25$ ; Group with  $8^\circ$ ; Target and Interference Both Offset Criteria

creased dramatically, without significant performance loss for SDMA systems.

## REFERENCES

- [1] Christof Farsakh and Josef A. Nossek, "Application of Space Division Multiple Access to Mobile Radio," *IEEE PIMRC*, vol. 2, pp. 736-739, September. 1994.
- [2] Christof Farsakh and Josef A. Nossek, "On The Mobile Radio Capacity Increase Through SDMA," *IEEE International Zurich Seminar on Broadband Comm.*, pp. 293-297, February. 1998.

- [3] Per Zetterberg and Bjorn Ottersten, "The Spectrum Efficiency of a Base Station Antenna Array System for Special Selective Transmission," *IEEE Transactions On Vehicular Technology*, vol. 44, no. 3, pp. 651-660, August. 1995.
- [4] K. I. Pedersen, P. E. Mogensen, B. H. Fleury, "Spatial Channel Characteristics in outdoor environments and their Impact on BS Antenna System Performance," *IEEE VTC*, vol. 2, pp.719-723, August. 1998.
- [5] Christof Farsakh and Josef A. Nossek, "Spatial Covariance Based Downlink Beamforming in an SDMA Mobile Radio System," *IEEE Transactions on Communications*, vol. 46, no. 11, pp.1497-1506, November. 1998.
- [6] Klaus Hugl, Juha Laurila and Ernst Bonek, "Downlink Performance of Adaptive Antennas With Null Broadening," *IEEE VTC*, vol. 1, pp.872-876, September. 1999.
- [7] Klaus Hugl, Juha Laurila and Ernst Bonek, "Downlink Performance for Frequency Division Duplex Systems," *IEEE Globecom*, vol. 4, pp.2097-2101, December. 1999.
- [8] Jaume Riba, Jason Goldberg and Gregori Vazquez, "Robust Beamforming for Interference Rejection in Mobile Communications," *IEEE Transactions on Signal Processing*, vol. 45, no. 1, pp.271-275, January. 1997.
- [9] Michael Tangemann, "Near Far Effects in Adaptive SDMA Systems," *IEEE PIMRC*, vol. 3, pp.1293-1297, September. 1995.

# ON THE USE OF CYCLOSTATIONARY FILTERS TO TRANSMIT INFORMATION

Alban DUVERDIER\*, Bernard LACAZE\*\* and Jean-Yves TOURNERET\*\*

\* CNES, 18 av. Belin, BPI 2012, 31401 Toulouse Cedex 4, France

\*\* ENSEEIHT/SIC, 2, rue Camichel BP7122, 31071 Toulouse Cedex 7, France

tel: +33 (0)5 61 28 31 79 / fax: +33 (0)5 61 28 26 13

email: Alban.Duverdier@cnes.fr

## ABSTRACT

Linear periodic time-varying filters are often introduced today in telecommunication. They spread the spectrum and can be used for scrambling, multi-user access or channel modeling. Recently, the authors have defined linear cyclostationary filters. In particular, this generalization has permitted to take into account the random parameters of a transmission channel. This paper defines a new case of linear cyclostationary filter where information is included into the filter.

We first recall the definition of linear periodic and linear cyclostationary filters. The paper presents then particular cases of these filters based on clock change. Thus, we introduce modulated periodic clock change. This filter can be used to transmit simultaneously an analog and a digital signal. We present the reconstruction method of the initial signals. We obtain reconstruction results in the case of the simultaneous transmission of an analog and a binary information.

## 1. INTRODUCTION

In telecommunications, signals subjected to a linear periodic filter [1] [2] are often encountered. Thus, this filter spread the spectrum and can correspond to a scrambling system [3], a multi-user access method [4] or a transmission channel modeling [5]. Recently, it was shown that they can be generalized in linear cyclostationary filters [6].

In the first section, we recall some definitions. In particular, we present the definition of linear cyclostationary filter. We introduce then a new filter called modulated periodic clock change. It permits to transmit simultaneously an analog and a digital signal. We present the reconstruction of the input signals. Finally, we apply the obtained reconstruction results to the transmission of an analog and a binary information.

## 2. DEFINITIONS

### 2.1. Stationary and cyclostationary processes

Let  $A = \{A(t), t \in \mathbf{R}\}$  be an harmonisable zero mean and mean square continuous process.  $A$  admits a Cramér-Loève representation  $\Theta_A(\omega)$  [7] such that:

$$A(t) = \int_{-\infty}^{+\infty} e^{i\omega t} d\Theta_A(\omega) \quad (1)$$

We note  $m_A(t)$  and  $R_A(t, \tau)$  the mean and autocorrelation function of  $A$  given by:

$$m_A(t) = E[A(t)] \quad (2)$$

$$R_A(t, \tau) = E[A(t + \tau/2)A^*(t - \tau/2)] \quad (3)$$

The power spectrum of  $A$ ,  $S_{A t}(\omega)$ , is defined by:

$$R_A(t, \tau) = \int_{-\infty}^{+\infty} e^{i\omega\tau} dS_{A t}(\omega) \quad (4)$$

$A$  is said to be stationary if and only if  $m_A(t)$  and  $R_A(t, \tau)$  are independent of  $t$ .  $dS_{A t}(\omega)$  is then independent of  $t$ .

$A$  is said to be cyclostationary if and only if  $m_A(t)$  and  $R_A(t, \tau)$  are periodic in  $t$  of period  $T = 2\pi/\omega_0$  [8].  $dS_{A t}(\omega)$  is then periodic in  $t$ . We suppose that it admits the Fourier series decomposition such that:

$$dS_{A t}(\omega) = \sum_{l=-\infty}^{+\infty} e^{il\omega_0 t} dS_A^l(\omega) \quad (5)$$

### 2.2. Linear time-invariant and periodic time-varying filters

Let  $\tilde{h}$  be a linear time-varying filter of frequency response  $h_t(\omega)$ . Its response to the stationary process  $Z$  is the process  $X$  defined by:

$$X(t) = \int_{-\infty}^{+\infty} e^{i\omega t} h_t(\omega) d\Theta_Z(\omega) \quad (6)$$



$\tilde{h}$  is a linear time-invariant filter if and only if  $h_t(\omega)$  is independent of the time.

$\tilde{h}$  is a linear periodic time-varying filter if and only if  $h_t(\omega)$  is periodic in time of period  $T$  [1]. We suppose that it admits the Fourier series decomposition such that:

$$h_t(\omega) = \sum_{l=-\infty}^{+\infty} e^{i\omega_0 t} h^l(\omega) \quad (7)$$

### 2.3. Linear stationary and cyclostationary filters

The linear random time-varying filter is a generalization of the linear time-varying filter previously defined [9]. Let  $\{H^\omega\}_{\omega \in \mathbb{R}}$  be a complex random processes family, where, for any  $\omega$ ,  $H^\omega = \{H_t(\omega), t \in \mathbb{R}\}$  is a complex continuous random process. We note  $\chi_t(\omega)$  the mean and  $\varphi_{t,\tau}(\omega, \gamma)$  the intercorrelation function of the  $\{H^\omega\}_{\omega \in \mathbb{R}}$ .  $\chi_t(\omega)$  and  $\varphi_{t,\tau}(\omega, \gamma)$  are given by:

$$\chi_t(\omega) = E[H_t(\omega)] \quad (8)$$

$$\varphi_{t,\tau}(\omega, \gamma) = E[H_{t+\frac{\gamma}{2}}(\omega + \frac{\gamma}{2}) H_{t-\frac{\gamma}{2}}^*(\omega - \frac{\gamma}{2})] \quad (9)$$

Let  $\tilde{h}$  be a linear random filter of frequency response  $H_t(\omega)$ . Its response to the stationary process  $Z$  is the process  $X$  defined by:

$$X(t) = \int_{-\infty}^{+\infty} e^{i\omega t} H_t(\omega) d\Theta_Z(\omega) \quad (10)$$

Thus, each linear filter can be seen as a particular case of linear random filter, where  $H^\omega$  is a degenerated random variable.

A linear random filter  $\tilde{h}$  is said to be stationary if and only if the processes  $\{H^\omega\}_{\omega \in \mathbb{R}}$  are jointly stationary. It means that the mean and the intercorrelation function of the  $\{H^\omega\}_{\omega \in \mathbb{R}}$  are independent of the time.

Recently, the authors have generalized this definition [6]. We call  $\tilde{h}$  a linear cyclostationary filter if and only if the processes  $\{H^\omega\}_{\omega \in \mathbb{R}}$  are jointly cyclostationary. It corresponds to the case where the mean and the intercorrelation function of the  $\{H^\omega\}_{\omega \in \mathbb{R}}$  are periodic in time of period  $T$ .

## 3. CLOCK CHANGE

### 3.1. Periodic clock change

The response  $X$  of a stationary process  $Z$  subjected to a periodic clock change [3]  $\tilde{h}$  is defined by:

$$X(t) = g(t)Z[t - f(t)] \quad (11)$$

where  $f(t)$  and  $g(t)$  are real measurable functions,  $T = 2\pi/\omega_0$  periodic. In equation (11),  $f(t)$  is a timing jitter and

$g(t)$  corresponds to an amplitude modulation. It is easy to see that a periodic clock change is a particular case of linear periodic filter and that its frequency response is given by:

$$h_t(\omega) = g(t)e^{-i\omega f(t)} \quad (12)$$

Periodic clock changes can be implemented easily. They appear also often in spread spectrum applications that use linear periodic filters, such as scrambling [3] and multi-user access [4].

### 3.2. Reconstruction of the input signal

Figure 1 depicts the reconstruction chain of a signal submitted to a periodic clock change.

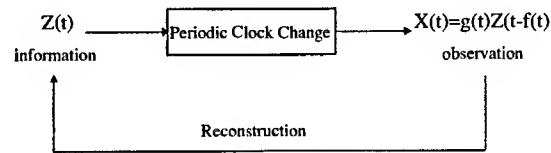


Figure 1: Reconstruction chain of a signal submitted to a periodic clock change

The reconstruction of a process subjected to a periodic clock change is a particular case of reconstruction of a process subjected to a linear periodic filter. Equations (6) and (7) show that the response  $X$  of the stationary process  $Z$  subjected to a linear periodic filter  $\tilde{h}$  admits the following spectral representation:

$$d\Theta_X(\omega) = \sum_{k=-\infty}^{+\infty} \psi_k(\omega - k\omega_0) d\Theta_Z(\omega - k\omega_0) \quad (13)$$

When the spectral support of  $Z$  is included in  $[-\omega_0/2, \omega_0/2[$ ,  $Z$  can then be reconstructed by:

$$\forall \omega \in [-\omega_0/2, \omega_0/2[, \forall k \in \Delta, d\Theta_Z(\omega) = \psi_k^{-1}(\omega) d\Theta_X(\omega + k\omega_0) \quad (14)$$

where  $\Delta$  is the integer set such that the functions  $\{\psi_k(\omega)\}_{k \in \Delta}$  are different from zero on the spectral support of  $Z$ . Multiple redundant reconstructions of  $Z$  can also be obtained by a frequency downconversion followed by a lowpass filtering on  $[-\omega_0/2, \omega_0/2[$ .

### 3.3. Modulated periodic clock change

The paper proposes a new clock change scheme that permits to transmit simultaneously an analog and a digital information. This spread spectrum technique is a generalization of

the classic periodic clock change. It can be useful for example to scramble video with analog image and digital sound. It is called modulated periodic clock change.

The response  $X$  of a stationary process  $Z$  subjected to such a clock change  $\tilde{h}$  is defined by:

$$X(t) = g(t)Z[t - M(t)f(t)] \quad (15)$$

where  $f(t)$  and  $g(t)$  are defined as in (11) and  $M = \{M(t), t \in \mathbb{R}\}$  is a stationary process independent of  $Z$ . Figure 2 depicts the obtained transmission chain.

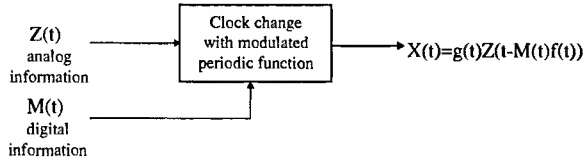


Figure 2: Transmission chain of a signal submitted to a modulated periodic clock change

It is easy to see that  $Z$  is then subjected to a cyclostationary filter of frequency response given by:

$$H_t(\omega) = g(t)e^{-i\omega M(t)f(t)} \quad (16)$$

In general, the reconstruction of  $Z(t)$  can be obtained by a sub-optimal solution [6]. Nevertheless, perfect reconstruction is possible when  $M$  is a Bernoulli variable that is equal to  $-1$  or  $+1$ .

In this case, equation (14) becomes:

$$\forall \omega \in [-\omega_0/2, \omega_0/2], \forall k \in \Delta, d\Theta_Z(\omega) = \psi_k^{-1}(M\omega) d\Theta_X(\omega + k\omega_0) \quad (17)$$

Let  $k_1$  and  $k_2$  be two values of  $k$ . Equation (17) implies that:

$$\forall \omega \in [-\omega_0/2, \omega_0/2], \psi_{k_1}^{-1}(M\omega) d\Theta_X(\omega + k_1\omega_0) = \psi_{k_2}^{-1}(M\omega) d\Theta_X(\omega + k_2\omega_0) \quad (18)$$

This equality allows the identification of  $M$  whenever  $\psi_{k_1}(\omega)$  and  $\psi_{k_2}(\omega)$  are not simultaneously even functions. Knowing  $M$ ,  $Z(t)$  can be perfectly reconstructed using (17).

This method can then be used for any binary signal  $M(t)$  whose sampling rate is much larger than  $T$ . It could be also generalized to any digital signal  $M(t)$ .

## 4. APPLICATION

### 4.1. Simultaneous transmission of an analog and a binary information

In the following simulations, a modulated periodic clock change is used to transmit simultaneously an analog signal

$Z(t)$  band-limited on  $[-\omega_0/2, \omega_0/2]$  and an N.R.Z. signal  $M(t)$ .  $f(t)$  and  $g(t)$  are given by:

$$f(t) = -\alpha \sin \omega_0 t \quad \text{and} \quad g(t) = 1 \quad (19)$$

Figure 3 depicts the analog signal at input of the clock change.

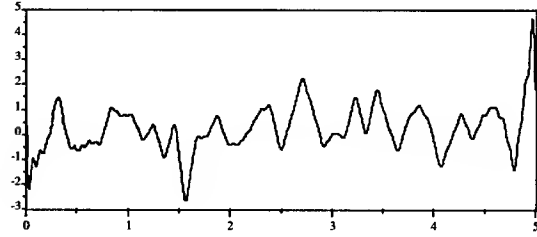


Figure 3: Initial analog signal

The binary signal is presented by Figure 4.

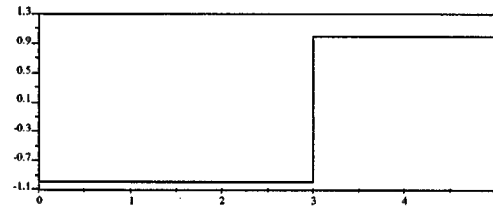


Figure 4: Initial binary signal

The signal observed at the output of the clock change is represented in Figure 5 for  $\alpha = 0.104$ ,  $T = 0.0347ms$  and a bit rate of  $1kb/s$ .

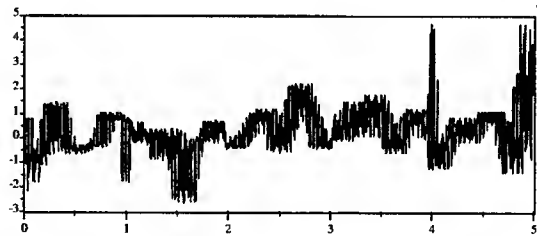


Figure 5: Observed signal

### 4.2. Reconstruction of the analog information

We have seen that  $Z(t)$  has to be reconstructed while  $M(t)$  is constant. As  $M(t)$  is a binary signal, the reconstruction

functions of  $Z(t)$  are given during each bit length by:

$$\psi_k(M\omega) = J_k(M\alpha\omega) \quad (20)$$

where  $J_k(\omega)$  is the  $k$ 'th order Bessel function and  $M$  is the value of  $M(t)$  that is equal to  $+1$  or  $-1$ . The reconstruction of  $Z(t)$  does not depend of  $M$  when  $k$  is even. It can then be obtained directly around any even  $k$ . Figure 6 compares the initial signal to the reconstruction obtained for  $k = 0$ . The analog information is well reconstructed.

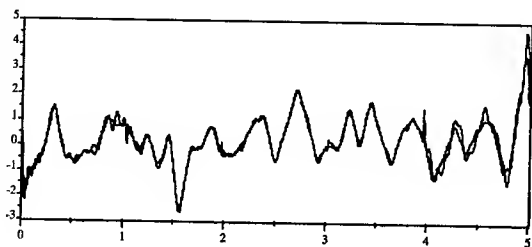


Figure 6: Reconstructed analog signal for  $k = 0$

#### 4.3. Reconstruction of the binary information

As we know a correct reconstruction of  $Z(t)$  for  $k$  even, the reconstructions obtained for  $k$  odd will allow to know when  $M(t)$  is correctly identified. Figure 7 and 8 compare the initial signal to the reconstruction for  $k = 1$ , when  $M(t)$  is supposed always equal to  $+1$  and when  $M(t)$  is correctly identified.

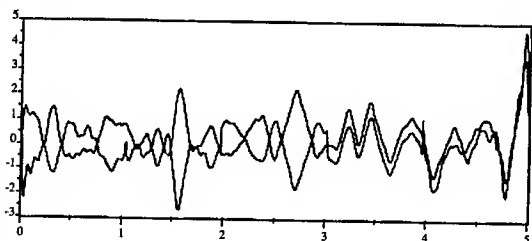


Figure 7: Reconstructed analog signal for  $k = 1$  with  $M(t)$  not correctly identified

The block diagram of Figure 9 shows a scheme which allows to reconstruct  $Z(t)$  and to recover the values of  $M(t)$  assuming perfect timing of the corresponding bit stream.

#### 5. CONCLUSION

In this paper, we recalled the definition of a linear periodic filter and of a linear cyclostationary filter. We presented a new filter called modulated periodic clock change. We

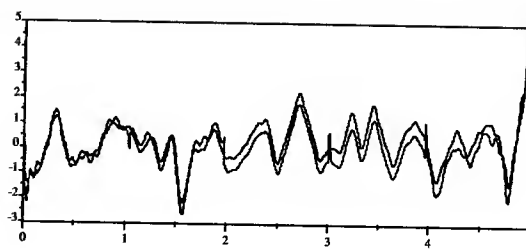


Figure 8: Reconstructed analog signal for  $k = 1$  with  $M(t)$  correctly identified

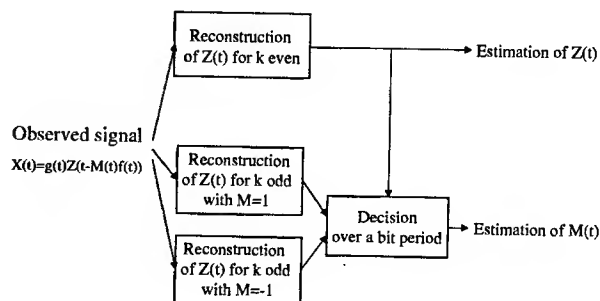


Figure 9: Scheme for the estimation of  $Z(t)$  and  $M(t)$

proposed a reconstruction method of the signals transmitted by this filter. It was applied successfully to the simultaneous transmission of an analog and a binary signal.

#### 6. REFERENCES

- [1] L.E. Franks, "Polyperiodic Linear Filtering" in *Cyclostationarity in Communications and Signal Processing*, William A. Gardner (eds.), IEEE Press, 1993
- [2] D. MacLernon, "Inter-relationships between different structures for periodic systems", *EUSIPCO*, 1998
- [3] A. Duverdier and B. Lacaze, "Time-varying reconstruction of stationary processes subjected to analogue periodic scrambling", *ICASSP*, 1997
- [4] A. Duverdier and B. Lacaze, "Transmission of two users by means of periodic clock changes", *ICASSP*, 1998
- [5] R.G. Gallager, *Information Theory and Reliable Communication*, Wiley, 1968
- [6] A. Duverdier, B. Lacaze and D. Roviras, "Introduction of linear cyclostationary filters to model time-variant channels", *GLOBECOM*, 1999
- [7] H. Cramer and M.R. Leadbetter, *Stationary and Related Stochastic Processes*, Wiley, 1967

- [8] W.A. Gardner and L.E. Franks, "Characterization of cyclostationary random signal processes", *IEEE Trans. Inform. Theory*, pp. 4-14, 1975
- [9] P.A. Bello, "Characterization of randomly time variant linear channels", *IEEE Trans. Comm.*, pp. 360-393, 1963

# NON-PARAMETRIC TRELLIS EQUALIZATION IN THE PRESENCE OF NON-GAUSSIAN INTERFERENCE

Carlo Luschi\*, Bernard Mulgrew

\* Bell Laboratories, Lucent Technologies

Unit 1, Pagoda Park, Westmead Drive, Swindon SN5 7YT, United Kingdom

Dept of Electronics and Electrical Engineering, University of Edinburgh  
The King's Buildings, Mayfield Road, Edinburgh EH9 3JL, United Kingdom

## ABSTRACT

We consider the problem of equalization of the frequency selective mobile radio channel in the presence of co-channel interference (CCI). Conventional trellis equalizers treat the sum of noise and interference as additive white Gaussian noise, while CCI is generally a colored non-Gaussian process. We propose a *non-parametric* approach based on the estimation of the probability density function of the noise-plus-interference. Given the availability of a limited volume of data, the density is estimated by kernel smoothing techniques. Due to the temporal color of the CCI, the use of a whitening filter is also addressed. Simulation results are given for the GSM system, showing a significant performance improvement with respect to the equalizer based on the Gaussian assumption.

## 1. INTRODUCTION

Time-division multiple access (TDMA) mobile radio systems like GSM are affected by co-channel interference (CCI) and intersymbol interference (ISI) due to multipath propagation. Channel equalizers commonly employed in practical GSM receivers perform maximum likelihood (ML) [1] or maximum *a posteriori* probability (MAP) [3] data estimation on the ISI trellis. ML sequence estimation using the Viterbi algorithm [2] is well known as the optimum detection technique for signals corrupted by finite-length ISI and additive white Gaussian noise (AWGN), in the sense that it minimizes the probability of a sequence error. The symbol-by-symbol MAP algorithm, proposed over two decades ago by Bahl *et al.* [3] for decoding of convolutional codes, has recently received renewed interest as a *soft-in/soft out* decoder for iterative decoding of parallel or serially concatenated codes [4]. As a trellis equalizer, the MAP algorithm is optimum in the sense that it minimizes the probability of symbol error. In receivers employing the concatenation of an equalizer and a channel decoder, the performance is improved by soft-decision decoding and iterative equalization and decoding [5]. In this respect, the MAP algorithm has the advantage of intrinsically providing optimal *a posteriori* probability as a soft-output value.

In this paper, we consider the problem of equalization of the mobile radio channel in the case of single channel reception. The optimum trellis equalizer in the presence

of ISI, CCI, and AWGN is based on joint detection of the co-channel signals [7]. Although joint ML and joint MAP detection are optimal, they can be prohibitively expensive since the complexity increases exponentially with the sum of the channel lengths of the desired and CCI signals. In addition, the estimation of the channel impulse response of all co-channel signals requires the knowledge of the training sequence of each interferer. On the other hand, conventional receivers employ a trellis equalizer which treats the sum of noise and interference as additive, white, Gaussian noise. In reality, the sum of noise and CCI is generally a colored non-Gaussian random process, and the above approach corresponds to a degradation of the error performance.

In order to correctly set the problem of trellis data estimation, a proper statistical characterization of the disturbance is required. To this purpose, we propose a *non-parametric* trellis equalizer, based on the estimation of the probability density function of the noise-plus-interference. Given the limited volume of training data, the work is based on the application of density estimation by *kernel smoothing*. The temporal color of the CCI is taken into account by a whitening filter.

## 2. MAP TRELLIS EQUALIZATION

### 2.1. System Model

Consider the received signal

$$r_k = \sum_{\ell=0}^{L-1} b_{k-\ell} h_{\ell}^{(k)} + n_k, \quad (1)$$

where  $b_k \in \{+1, -1\}$  are the transmitted symbols, the  $L$  complex tap-gains  $h_{\ell}^{(k)}$  represent the samples of the equivalent channel impulse response at time  $k$ , and  $n_k = y_k + w_k$  indicates the sum of co-channel interference and thermal noise. In the case of the GSM system, we consider the linearized model of the GMSK signal [8], where  $h_{\ell}^{(k)}$  are the taps of the equivalent discrete-time channel produced by derotation of the received signal [9]. The GSM signal has almost zero excess bandwidth, and we assume that sufficient statistics for data estimation can be obtained by symbol-rate sampling at the output of a fixed front-end filter. The

analysis can be extended to include the case of non-zero excess bandwidth by introducing oversampling and fractionally spaced trellis equalization.

In this Section, we consider the CCI samples as independent complex non-Gaussian random variables. The discrete-time process  $y'_k$  is generally colored, even if the delay spread in a typical interference-limited environment is usually relatively small. At high signal-to-noise-ratios (SNRs) a suitable temporal prewhitening is assumed to produce approximately independent non-Gaussian disturbance. The validity of this assumption will be discussed in Section 3.

## 2.2. Symbol-by-Symbol MAP Algorithm for Finite-Length ISI and Additive Independent Disturbance

Suppose that the symbols  $b_k$  are transmitted in finite blocks of length  $N$ . Assuming the knowledge of the channel impulse response, a *soft-output* symbol-by-symbol MAP equalizer computes the *a posteriori* log-likelihood ratio

$$L(b_k|r_0, \dots, r_{N-1}) \triangleq \log \frac{\Pr(b_k = +1|r_0, \dots, r_{N-1})}{\Pr(b_k = -1|r_0, \dots, r_{N-1})}, \quad (2)$$

with  $0 \leq k \leq N-1$ . Let  $\mu_k \triangleq (b_{k-1}, \dots, b_{k-L+1})$  denote the generic ISI state at time  $k$ , and  $S(b_k)$  the set of states corresponding to the transmitted symbol  $b_k$ . Indicating by  $\xi_k$  the transition from the state  $\mu_k$  to  $\mu_{k+1}$ , the MAP algorithm results in a *forward* and *backward* recursions with the transition metric  $\lambda(\xi_k)$ , coupled by a dual-maxima operation [3], [6]

$$L(b_k|r_0, \dots, r_{N-1}) = \max'_{\mu \in S(b_k=+1)} \Lambda(\mu_{k+1}) - \max'_{\mu \in S(b_k=-1)} \Lambda(\mu_{k+1}) \quad (3)$$

$$\Lambda(\mu_{k+1}) = \Lambda^f(\mu_k) - \lambda(\xi_k) + \Lambda^b(\mu_{k+1}), \quad (4)$$

where  $\Lambda(\mu_k)$  is the overall accumulated metric for the state  $\mu_k$ ,  $\Lambda^f$  and  $\Lambda^b$  are the accumulated metrics in the forward and backward recursions, and  $\max'\{x, y\} \triangleq \max\{x, y\} + \log(1 + e^{-|x-y|})$  [6]. The metric increment  $\lambda(\xi_k)$  results

$$\lambda(\xi_k) = -\log p(r_k|b_k, \dots, b_{k-L+1}) - \log \Pr(b_k), \quad (5)$$

where  $p(r_k|b_k, \dots, b_{k-L+1}) = p_n(r_k - \sum_{\ell=0}^{L-1} b_{k-\ell} h_\ell^{(k)})$ . In the case where  $n_k$  is modelled as AWGN, the quantity  $-\log p(r_k|b_k, \dots, b_{k-L+1})$  in (5) produces the Euclidean distance metric. When no *a priori* information is available about the transmitted bit  $b_k$ , the term  $-\log \Pr(b_k)$  in (5) has no effect and can be omitted from the calculation. On the contrary, if the equalizer receives some *a priori* information the above term has a fundamental role in deriving a *soft-in/soft-out* MAP equalizer [4], [5].

Observe that the above derivation relies on the assumption of known channel. In practice, the channel response is usually estimated using a known training sequence at the equalizer start-up.

## 3. TRELLIS EQUALIZATION BY NON-PARAMETRIC DENSITY ESTIMATION

### 3.1. Density Estimation by Kernel Smoothing

An example of the density function of the noise plus CCI samples  $n_k$  for the case of the GSM channel is shown in

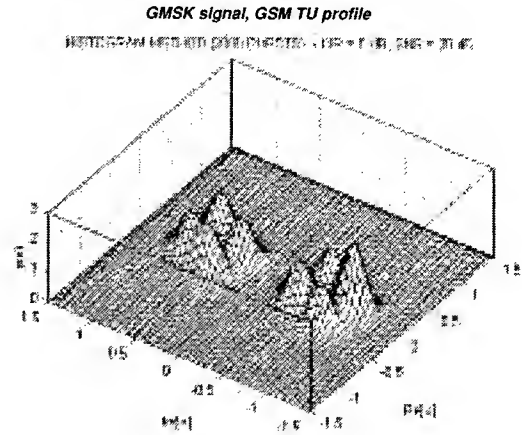


Figure 1: Example of the density function of CCI (derotated GMSK signal) plus AWGN for a GSM receiver.

Figure 1. The plot has been obtained by a histogram of the data in 2000 bursts, considering one dominant interferer under stationary propagation conditions. From Figure 1, it is apparent that the disturbance can not be realistically modelled as a Gaussian random variable.

#### 3.1.1. Parzen Estimator

An estimate of the probability density function of a complex random variable  $X$  can be built from a set of data  $X_i, i = 1, \dots, n$ , by means of a *smoothing function* or *kernel function*  $K(x, X_i)$  (see [11] and references therein). In the method proposed by Parzen [10], an estimate of the unknown density is given by

$$\hat{p}_n(x) = \frac{1}{n} \cdot \sum_{i=1}^n K(x, X_i). \quad (6)$$

A possible choice for the function  $K(x, X_i)$  among those satisfying the conditions for (asymptotic) unbiasedness and consistency of the estimator [10] is the Gaussian kernel of fixed width  $\sigma_0$

$$K(x, X_i) = \frac{1}{2\pi\sigma_0^2} e^{-|x-X_i|^2/2\sigma_0^2}. \quad (7)$$

#### 3.1.2. Transition Metrics for Non-Parametric Trellis Equalization

In the case of a Bayesian trellis equalizer, the random variable  $X$  represents one realization of the process of noise-plus-interference corresponding to a given received burst. Consider the received signal (1), and assume that the channel is approximately constant within the burst duration. Then, once the channel taps  $h_\ell$  are estimated using the  $M$  training symbols  $\hat{b}_i$ , they can be used to derive the set of observations  $X_i, i = 1, \dots, n = M - L$  of the random disturbance  $X$  according to  $X_i = \hat{n}_i = r_i - \sum_{\ell=0}^{L-1} \hat{b}_{i-\ell} \hat{h}_\ell$ ,

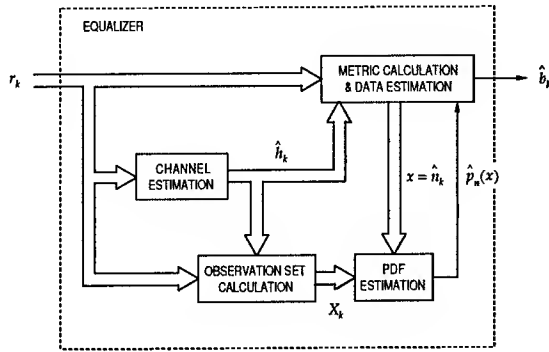


Figure 2: Block diagram of the non-parametric trellis equalizer.

where hat denotes the estimated value. At this point we recall that the transition metric (5) of the optimum symbol-by-symbol MAP algorithm results  $\lambda(\xi_k) = -\log \hat{p}_n(r_k - \sum_{\ell=0}^{L'-1} b_{k-\ell} h_\ell) - \log \Pr(b_k)$ . Therefore, using (6) and (7) one can directly estimate the quantity  $\log \hat{p}_n(x)$  for  $x = \hat{n}_k = r_k - \sum_{\ell=0}^{L'-1} b_{k-\ell} \hat{h}_\ell$ , and obtain

$$\lambda(\xi_k) = -\log \hat{p}_n(x) - \log \Pr(b_k). \quad (8)$$

The block diagram of the resulting equalizer is shown in Figure 2. From the implementation point of view, the density  $\log \hat{p}_n(x)$  at time  $k$  can be computed separately for each trellis branch. Alternatively, it can be precomputed for a finite number of values  $x$ , and stored in a look-up-table before starting the trellis processing.

We emphasize the fact that the above technique deals with the statistical model of a random variable, obtained as the realization of the noise-plus-interference process at a given time instant. It is worth noting that, with a proper adaptive procedure, the approach can be extended to those cases where the CCI impulse response cannot be considered approximately constant within the burst.

### 3.2. Probability Density Function of the Noise-plus-Interference

The analytical expression of the actual density function of noise-plus-interference can be carried out if we assume a (unknown) deterministic finite-state machine model for the co-channel signal. Consider the received signal (1). The sum of noise and CCI at time  $k$  can be expressed as

$$n_k = y'_k + w_k = \sum_{\ell=0}^{L'-1} b'_{k-\ell} h'_\ell + w_k, \quad (9)$$

where  $b'_k \in \{+1, -1\}$  are the co-channel symbols,  $h'_\ell$ ,  $0 \leq \ell \leq L'-1$  denote the taps of the co-channel impulse response, and  $w_k$  is white Gaussian noise with zero mean

and variance  $2\sigma^2$ , which we assume independent of  $y'_k$ . If the co-channel taps  $h'_\ell$  at time  $k$  are regarded as an unknown, but deterministic mapping from  $(b'_k, \dots, b'_{k-L'+1})$  to  $y'_k$ , the distribution of  $n_k$  can be derived from those of  $b'_k$  and  $w_k$ . Given a generic binary quantity  $\beta$ , we define

$$\eta_i = \eta_{i,1} + j\eta_{i,2} \triangleq \sum_{\ell=0}^{L'-1} \beta_{i,\ell} h'_\ell, \quad 0 \leq i \leq 2^{L'} - 1, \quad (10)$$

where  $\beta_i = \{\beta_{i,\ell}\}_{\ell=0}^{L'-1}$  denotes one of the  $2^{L'}$  distinct sequences of elements  $\beta_{i,\ell} \in \{+1, -1\}$ . Then, it is possible to show that the expression of the density of  $n_k$  results

$$p_n(x) = \frac{1}{2^{L'}} \sum_{i=1}^{2^{L'}} p_w(x - \eta_i), \quad (11)$$

where  $p_w(x)$  is the complex Gaussian density with variance  $2\sigma^2$ . From (11), the density of the interference-plus-noise is given by a number of symmetric Gaussian kernels, which centers are the points of the hypothetical scatter diagram obtained in the absence of thermal noise. Comparison of (11) and (6) reveals the strong connection between the structure of the Parzen estimator and the true density. In particular, for  $\sigma^2 \rightarrow 0$ , the observations  $X_i$  in (6) correspond to the points of the complex plane defined by (10), with the binary parameters  $\beta_{i,\ell}$  replaced by the co-channel symbols  $b'_{k-\ell}$ . Therefore, the estimator defined by (6) and (7) will approach the true density (11) as soon as the dimension of the training data is large enough to represent the  $2^{L'}$  equiprobable sequences  $\beta_i = \{\beta_{i,\ell}\}_{\ell=0}^{L'-1}$ .

### 3.3. Doubling the Size of the Training Set

We observe that in (10) for each index  $i = i'$  corresponding to the binary sequence  $\beta_{i'} = \{\beta_{i',\ell}\}_{\ell=0}^{L'-1}$  there is an index  $i = i''$  with  $\beta_{i''} = \{-\beta_{i',\ell}\}_{\ell=0}^{L'-1} = -\beta_{i'}$ . This means that for each  $i'$  there is an  $i''$  such that  $\eta_{i'} = -\eta_{i''}$ . Exchanging each pair of indexes  $i'$  and  $i''$  in the sum (11) and taking into account the symmetry of the Gaussian density  $p_w(x)$  gives  $p_n(-x) = (1/2^{L'}) \sum_{i=1}^{2^{L'}} p_w(-x + \eta_i) = p_n(x)$ . The importance of this result comes from the fact that it allows to double the available volume of data in the density estimator (6). In fact it implies that, if  $\{X_i\}$  are values assumed by the random variable  $n_k$ , then the set  $\{-X_i\}$  contains values assumed by  $n_k$  with the same probability. Therefore, together with each outcome  $X_i$  we can additionally consider  $-X_i$  as if it was the result of a parallel experiment. This leads to the enlarged data set  $\{X_i, -X_i\}$ .

### 3.4. Choice of the Smoothing Parameter

An optimal kernel width for the fixed-width density estimator (6) can be determined through the minimization of the mean integrated square error (MISE) [11]. In the case of the Gaussian kernel (7) used to estimate the complex Gaussian density with variance  $2\sigma^2$ , we have  $\sigma_{0(opt)} = (1/n)^{1/6} \sigma$  [11]. For the density of the noise-plus-interference, using (11) and applying Cauchy's inequality we find

$$\sigma_{0(opt)} \geq (1/n)^{1/6} \sigma. \quad (12)$$

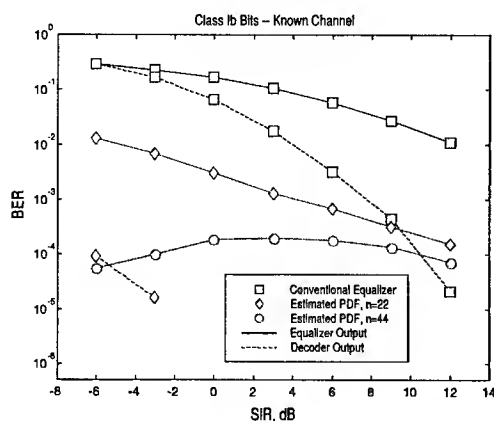


Figure 3: Error performance in the case of known channel. GSM TU0 profile, SNR = 30 dB. Density estimator with fixed kernel width  $\sigma_0 = 0.05$ .

With a given volume  $n$  of training data, the kernel width can then be selected from the value of the noise variance  $\sigma^2$ . In a practical receiver, an estimate of  $\sigma^2$  can be derived by the training sequence, taking into account the estimated channel response and the measure of the received signal level.

### 3.5. Temporal Whitening

The MAP equalizer with branch metric (8) is based on the assumption that the samples  $n_k$  are independent. Given the temporal color of the CCI, a whitening filter of the disturbance is needed before the trellis processor. We point out that a linear prediction-error (LPE) filter will ideally produce uncorrelated CCI-plus-noise samples, but this does not necessarily imply independence, since the process continues in general to be non-Gaussian. In addition, a whitening filter for the disturbance will inevitably increase the channel memory for the desired signal. And if we do not want to increase the number of states of the equalizer, the number of taps of the filter has to be kept small. However, the delay spread of the typical GSM urban channel is usually lower than 4 symbol intervals. Moreover, reducing the correlation between the samples will certainly reduce their 'dependence'. Note that in some particular cases the whitened disturbance turns out to be actually independent. As an example, this happens when the variance of the thermal noise tends to zero and the co-channel is minimum-phase (in fact, in this case the ideal LPE filter inverts the co-channel).

## 4. SIMULATION RESULTS

The effectiveness of the strategy based on density estimation by kernel smoothing has been assessed by computer simulation for the case of a GSM receiver with single channel reception. The GMSK transmitted symbols are ob-

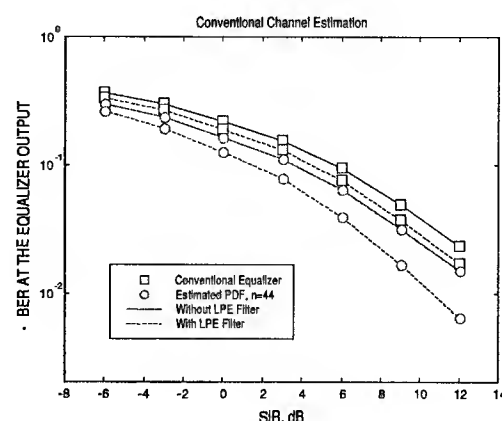


Figure 4: Error performance in the case of estimated channel. GSM TU0 profile, SNR = 30 dB. Density estimator with fixed kernel width  $\sigma_0 = 0.05$ .

tained from the source bits by rate 1/2 convolutional encoding and interleaving, according to the GSM specifications for the full-rate speech traffic channel. The simulator includes the multipath fading channel with the classical Doppler spectrum [14], CCI, and thermal noise. Ideal frequency hopping is implemented. One dominant co-channel interferer is assumed, characterized by an independent fading process and a random phase shift with respect to the signal of interest. In all the simulations SNR = 30 dB. At the receiver, the soft-output data produced by a 16-states MAP equalizer are deinterleaved and decoded by a convolutional channel decoder.

To establish the ultimate performance of the proposed equalizer, we first consider the ideal case of known channel and relative speed 0 Km/h. Figure 3 shows the bit-error rate (BER) performance with GSM typical urban area (TU) multipath profile for both co-channel signals. The MAP non-parametric equalizer is compared with the MAP trellis processor that assumes Gaussian disturbance. The figure also addresses the effect of doubling the data set for density estimation, as discussed in Section 3. The results indicate that the non-parametric equalizer offers a potential improvement of more than two orders of magnitude in terms of BER at the equalizer output. Figures 4 to 6 illustrate the receiver performance when the channel of the signal of interest is estimated from the training symbols. We also introduce an LPE filter for prewhitening of the colored disturbance. As discussed in Section 3, choosing the prediction order involves a trade-off between performance and complexity. In the figures, we use a 16-states trellis and a 2-taps LPE filter. Finally, we include the performance obtained by iterative channel estimation. In this case, after the equalization of the entire burst, the data decisions are fed back to produce an improved channel estimate, which is used in a second pass equalization.

The above simulation results refer to a synchronous



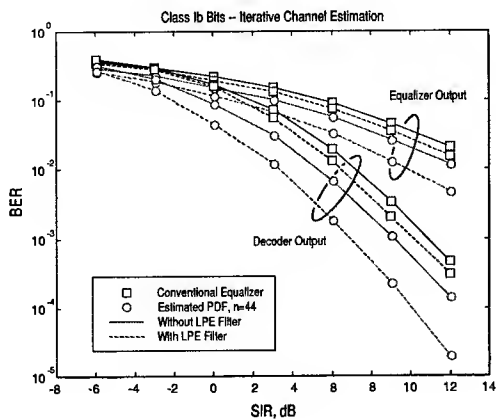


Figure 5: Error performance with iterative channel estimation. GSM TU0 profile, SNR = 30 dB. Density estimator with fixed kernel width  $\sigma_0 = 0.05$ .

interference scenario. Simulation with asynchronous CCI shows that the proposed equalizer still outperforms the conventional trellis processor. However, in those cases the proper approach consists in introducing an adaptation of the estimated density of the noise-plus-CCI.

## 5. CONCLUSIONS

A non-parametric trellis processor has been studied for channel equalization in the presence of non-Gaussian interference. In the case of the GSM system, the proposed approach based on density estimation by kernel smoothing provides a significant performance improvement with respect to the receiver that assumes Gaussian disturbance.

## REFERENCES

- [1] G. D. Forney, Jr., "Maximum likelihood sequence estimation of digital sequences in the presence of intersymbol interference," *IEEE Trans. Inform. Theory*, vol. IT-18, no. 3, pp. 363-378, May 1972.
- [2] G. D. Forney, Jr., "The Viterbi algorithm," *Proc. IEEE*, vol. 61, no. 3, pp. 268-278, Mar. 1973.
- [3] L. R. Bahl, J. Cocke, F. Jelinek, and J. Raviv, "Optimal decoding of linear codes for minimizing symbol error rate," *IEEE Trans. Inform. Theory*, vol. IT-20, pp. 284-287, Mar. 1974.
- [4] S. Benedetto, D. Divsalar, G. Montorsi, and F. Polara, "A soft-input soft-output APP module for iterative decoding of concatenated codes," *IEEE Commun. Letters*, vol. 1, no. 1, pp. 22-24, Jan. 1997.
- [5] G. Bauch, H. Khorram, and J. Hagenauer, "Iterative equalization and decoding in mobile communications

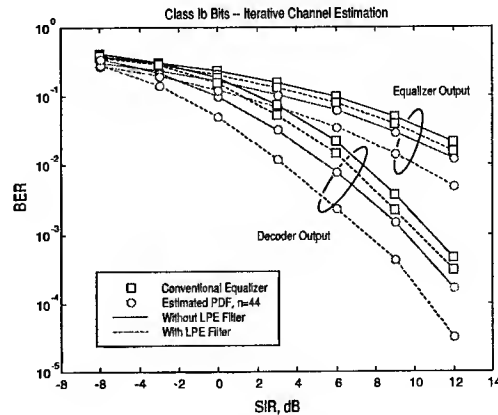


Figure 6: Error performance with iterative channel estimation. GSM TU50 profile, SNR = 30 dB. Density estimator with fixed kernel width  $\sigma_0 = 0.05$ .

systems," in *Proc. Eur. Pers. Mobile Commun. Conf.*, (Bonn, Germany), pp. 307-312, Oct. 1997.

- [6] A. J. Viterbi, "An intuitive justification and a simplified implementation of the MAP decoder for convolutional codes," *IEEE J. Select. Areas Commun.*, vol. 16, no. 2, pp. 260-264, Feb. 1998.
- [7] K. Giridhar, J. J. Shynk, A. Mathur, S. Chari, and R. P. Gooch, "Nonlinear techniques for the joint estimation of cochannel signals," *IEEE Trans. Commun.*, vol. 45, no. 4, pp. 473-484, Apr. 1997.
- [8] P. Laurent, "Exact and approximate construction of digital phase modulations by superposition of amplitude modulated pulses (AMP)," *IEEE Trans. Commun.*, vol. 34, no. 2, pp. 150-160, Feb. 1986.
- [9] A. Baier, "Derotation techniques in receivers for MSK-type CPM signals," in *Proc. Eusipco*, (Barcelona, Spain), Sept. 1990.
- [10] E. Parzen, "On estimation of a probability density function and mode," *Ann. Math. Statist.*, vol. 33, pp. 1065-1076, 1962.
- [11] A. W. Bowman and A. Azzalini, *Applied Smoothing Techniques for Data Analysis*. Oxford: Oxford University Press, 1997.
- [12] J.-N. Hwang, S.-R. Lay, and A. Lippman, "Nonparametric multivariate density estimation: A comparative study," *IEEE Trans. Signal Proc.*, vol. 42, no. 10, pp. 2795-2810, Oct. 1994.
- [13] C. Diamantini and A. Spalvieri, "Quantizing for minimum average misclassification risk," *IEEE Trans. Neural Networks*, vol. 9, no. 1, pp. 174-182, Jan. 1998.
- [14] J. G. Proakis, *Digital Communications*. New York: McGraw-Hill, 3rd ed., 1995.

# ANALYTICAL BLIND IDENTIFICATION OF A SISO COMMUNICATION CHANNEL

*Olivier GRELLIER and Pierre COMON*

Lab. I3S, Algorithmes-Euclide-B, 2000 route des Lucioles  
BP 121, F-06903 Sophia-Antipolis cedex, France  
grellier@i3s.unice.fr comon@unice.fr

## ABSTRACT

In this paper, a novel analytical blind identification algorithm is presented, based on the non-circular second-order statistics of the output. It is shown that the channel taps need to satisfy a polynomial system of degree 2, and that identification amounts to solving the system. We describe the algorithm able to solve this particular system entirely analytically. Computer results demonstrate its efficiency.

## 1. INTRODUCTION

Blind identification methods depend on the characteristics of the input sources. For example, it is known that a system can only be identified up to an all-pass filter when its input is Gaussian circular. Consequently, a particular attention has been paid to the non-Gaussian input cases. In those situations the phase information can be accessed using high-order statistics of the observations, and in the SISO case the system is identified up to a scalar factor only. This has been studied in numerous papers among which one can cite the works of Shalvi-Weinstein [5], Tugnait [7].

An interesting class of non-Gaussian signals is the discrete one, which appears in wireless communications. The discrete character has been used by few authors such as Li [3] or Yellin and Porat [8], who were the first interested in an algebraic solution. The studied signals have also non zero cyclostationary statistics, which allows identification using second-order statistics only [4] [6].

The novelty of our contribution is two-fold. First, non-circular second-order moments are used. Second, an algebraic solution to a class of polynomial systems, constructed from a block of data, is introduced. Our approach is described in the case of MSK modulations, approximating well the digital modulation utilized in the GSM standard. In addition, block methods are well matched to burst-mode communication systems.

## 2. MODEL, NOTATION, AND ASSUMPTIONS

Assume a finite sequence of input samples  $x(m)$  is fed into a Finite Impulse Response (FIR) linear system of length  $M$ . Denote  $y(n)$  the corresponding output sequence of length  $N$ , satisfying:

$$y(n) = \sum_{m=0}^{M-1} h(m) x(n-m) + w(n) \stackrel{\text{def}}{=} \mathbf{x}(n; M)^T \mathbf{h} + w(n)$$

Multidimensional variables are stored in column vectors and denoted by boldface letters; for instance,  $\mathbf{x}(n; M) = [x(n), \dots, x(n-M+1)]^T$ , by construction.

The input sequence is assumed to follow a discrete distribution, stemming from BPSK, MSK, or QPSK digital modulations, and the channel  $\mathbf{h}$  is supposed time-invariant during the observation.

The key statistical property used in this paper is that discrete signals are non-stationary at given orders. More precisely, for BPSK modulated signals :

$$\begin{aligned} E\{x(n)x(n-\ell)|x(0)\} &= x(0)^2\delta(\ell) \\ E\{x(n)x(n-\ell)^*\} &= \delta(\ell) \end{aligned}$$

for MSK signals :

$$\begin{aligned} E\{x(n)x(n-\ell)|x(0)\} &= (-1)^n x(0)^2\delta(\ell) \\ E\{x(n)x(n-\ell)^*|x(0)\} &= \delta(\ell) \end{aligned}$$

and for QPSK modulated signals:

$$\begin{aligned} E\{Re[x(n)] Re[x(n-\ell)]|x(0)\} &= Re[x(0)]^2\delta(\ell) \\ E\{Im[x(n)] Im[x(n-\ell)]|x(0)\} &= Im[x(0)]^2\delta(\ell) \\ E\{x(n)x(n-k)x(n-\ell)x(n-m)|x(0)\} &= x(0)^4\delta(k+\ell+m) \\ E\{x(n)x(n-\ell)^*\} &= \delta(\ell), \end{aligned}$$

and where  $\delta(\ell) \stackrel{\text{def}}{=} 1$  if  $\ell = 0$  and  $\delta(\ell) = 0$  elsewhere. Note the conditional expectation, exhibiting cyclostationarity in the non-circular moment of MSK inputs.

Based on these properties, it is possible to derive a set of polynomial equations that the channel must satisfy. In the MSK case, we obtain :

$$E[y(n)y(n-\ell)|x(0)] = x(0)^2 \sum_{m=0}^{M-1} (-1)^m h(m)h(m+\ell)$$

In the BPSK case, we have :

$$E[y(n)y(n-\ell)|x(0)] = x(0)^2 \sum_{m=0}^{M-1} h(m)h(m+\ell)$$

lastly in the QPSK case :

$$E[y(n)y(n-\ell_1)y(n-\ell_2)y(n-\ell_3)|x(0)] = x(0)^4 \sum_{m=0}^{M-1} h(m)h(m+\ell_1)h(m+\ell_2)h(m+\ell_3)$$

### 3. SOLVING THE POLYNOMIAL SYSTEM

#### 3.1. Example

In order to introduce in easy words our contribution, let's give a simple example. Let the input signal be MSK and the channel be real of length  $M = 3$ . Then non circular statistics yield:

$$\begin{cases} h(0)^2 - h(1)^2 + h(2)^2 &= f_1 \\ h(0)h(1) - h(1)h(2) &= f_2 \\ h(0)h(2) &= f_3 \end{cases} \quad (1)$$

whereas circular ones yield:

$$\begin{cases} h(0)^2 + h(1)^2 + h(2)^2 &= g_1 \\ h(0)h(1) + h(1)h(2) &= g_2 \\ h(0)h(2) &= f_3 \end{cases}$$

where  $f_i$  and  $g_i$  are given (they depend on statistics of observations  $y$ ). The grouping of those equations allows to obtain:

$$\begin{cases} h(0)^2 + h(2)^2 &= (f_1 + g_1)/2 \\ h(0)h(1) &= (f_2 + g_2)/2 \\ h(0)h(2) &= f_3 \end{cases}$$

Using the first and third equations, one gets:

$$\begin{aligned} (h(0) - ih(2))^2 &= h(0)^2 + h(2)^2 - 2ih(0)h(2) \\ &= (f_1 + g_1)/2 - 2if_3 \end{aligned}$$

This equation eventually allows to calculate  $h(0)$  and  $h(2)$  up to a sign, and then  $h(1)$ .

Thus we have been able to identify a real channel by using the **non-circular second order** statistics together with **circular second order** ones. The general algorithm that is described in this section computes the finite set of solutions of the polynomial system built on the non-circular second-order statistics only. In the next section, the choice of the channel estimation is discussed.

#### 3.2. Preliminaries

Consider the ring  $\mathcal{R} = \mathcal{C}[\xi]$  of polynomials in variables  $\xi \stackrel{\text{def}}{=} [h(0), h(1), \dots, h(M-1)]$  with coefficients in the complex field  $\mathcal{C}$ ; the dual space of  $\mathcal{R}$  is the set of linear forms from  $\mathcal{R}$  to  $\mathcal{C}$ , denoted  $\hat{\mathcal{R}}$ . The evaluation of a polynomial  $p$  at a point  $\zeta \in \mathcal{C}^M$ , denoted by  $1_\zeta : p \mapsto p(\zeta)$ , is the linear form which we are most interested in.

Given a polynomial  $a \in \mathcal{A}$ , define the multiplication operator by  $a$  as the mapping  $\mathcal{M}_a$  that associates  $q$  with  $aq$  :

$$\begin{aligned} \mathcal{M}_a : \mathcal{A} &\rightarrow \mathcal{A} \\ q &\mapsto qa \end{aligned} \quad (2)$$

The transposed operator,  $\mathcal{M}_a^T$ , is by definition the mapping from  $\mathcal{A}$  onto itself so that  $\langle q, \mathcal{M}_a^T \Lambda \rangle = \langle \mathcal{M}_a q, \Lambda \rangle = \langle aq, \Lambda \rangle$ ,  $\forall \Lambda \in \hat{\mathcal{A}}$ ,  $\forall q \in \mathcal{R}$  so that  $\mathcal{M}_a^T(\Lambda)(q) = \Lambda(qa)$ .

#### 3.3. Lemmas

Let  $\mathcal{P}$  be the subset  $\mathcal{R}$  of polynomials  $\{f_1, \dots, f_M\}$  of degree  $D$  and belonging to  $\mathcal{R}$ . Bézout's theorem [2, p.227] states that such a system

$$\mathcal{P} : \{f_m(\xi) = 0, 1 \leq m \leq M\} \quad (3)$$

where  $\xi \stackrel{\text{def}}{=} [\xi(0), \xi(1), \dots, \xi(M-1)]$ , has an infinity of solutions, or a number of solutions smaller or equal to  $D^M$ .

When the system has a finite number of solutions, one conventional way to compute them is to reduce the problem to an eigenvector computation, as shown by the following lemma.

**Lemma 3.1** *Linear forms  $1_\xi : p \mapsto p(\xi)$ , where  $\xi$  is any solution of  $\mathcal{P}$ , are the eigenvectors of all matrices  $(M_a^T)_{a \in \mathcal{A}}$ . The corresponding eigenvalues are  $a(\xi)$ .*

For a proof see [1].

Therefore, the computation of the multiplication matrix  $M_a$  appears as a key step in the proposed algorithm, since its eigen vectors allows to find the solutions of  $\mathcal{P}$ . Indeed, if we take for a basis of  $\mathcal{A}$ ,  $\mathcal{B} = \{1, h(0), h(1), \dots, h(0)h(1), \dots\}$ , the entries of the eigenvectors are equal to  $\{1, \xi(0), \xi(1), \dots, \xi(0)\xi(1), \dots\}$ , where  $\xi$  stands for any possible solutions of  $\mathcal{P}$ .

#### 3.4. Computation of matrix $M_a$

Matrix  $M_a$  can be directly computed from the Macaulay matrix associated with polynomials

$\{f_1, \dots, f_M\}$  [1]. These matrices are the extension of the so-called Sylvester matrices to multivariate polynomials.

However, if we take into account the relationships between the monomials introduced in the polynomial system  $\mathcal{P}$ , there exists a much simpler procedure to compute  $M_a$ .

In order to simplify the discussion, we will use the following example. Suppose that a channel of length  $M = 3$  is excited by a MSK input. System  $\mathcal{P}$  is then equal to that in (1). In this case, a generic basis that can solve this kind of system is given by :

$$\mathcal{B}_3 = \{1, h(0), h(1), h(2), h(0)h(1), h(0)h(2), h(1)h(2), h(0)h(1)h(2)\}$$

However, this basis cannot be used in our problem unless we first apply a change in the variables. Thus, the computation of the multiplication matrix can be split into 4 steps.

#### First step : Change in variables

Suppose we use the following change in variables :  $\xi = T\mathbf{h}$  then the system in  $\xi$  becomes :

$$A\mathcal{H} = \begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix} \quad (4)$$

where the entries of  $A$  are functions of the entries of  $T$ , and :

$$\mathcal{H} = [1, h(0), h(1), h(2), h(0)h(1), h(0)h(2), h(1)h(2), h(0)^2, h(1)^2, h(2)^2]$$

The matrix  $T$  must be chosen so that monomials  $h(0)^2$ ,  $h(1)^2$  and  $h(2)^2$  can be directly expressed as functions of the basis (see the next step).

#### Second step : Expression of the second degree monomials

Suppose we want to find the matrix associated with the multiplication by  $h(0)$ . The monomials that we have to express are :

Monomials of the basis		Monomials to be expressed
1		$h(0)$
$h(0)$		$h(0)^2$
$h(1)$		$h(0)h(1)$
$h(2)$		$h(0)h(2)$
$h(0)h(1)$	$\xrightarrow{\times h(0)}$	$h(0)^2h(1)$
$h(0)h(2)$		$h(0)^2h(2)$
$h(1)h(2)$		$h(0)h(1)h(2)$
$h(0)h(1)h(2)$		$h(0)^2h(1)h(2)$

Among these monomials, some are already in the basis. In our example, the monomials  $h(0)$ ,  $h(0)h(1)$ ,  $h(0)h(2)$  et  $h(0)h(1)h(2)$  are in the basis. The others monomials,  $h(0)^2$ ,  $h(0)^2h(1)$ ,  $h(0)^2h(2)$  and  $h(0)^2h(1)h(2)$ , have to be expressed using the polynomial system.

According to equation (4), monomials  $h(0)^2$ ,  $h(1)^2$  and  $h(2)^2$  can be expressed directly as a function of  $1$ ,  $h(0)$ ,  $h(1)$ ,  $h(2)$ ,  $h(0)h(1)$ ,  $h(0)h(2)$  and  $h(1)h(2)$ , provided that  $T$  is chosen correctly. In other words, monomials  $h(0)^2$ ,  $h(1)^2$  and  $h(2)^2$  can be expressed directly as a function of the basis using equation (4).

$$\begin{bmatrix} h(0)^2 \\ h(1)^2 \\ h(2)^2 \end{bmatrix} = B \begin{bmatrix} 1 \\ h(0) \\ h(1) \\ h(2) \\ h(0)h(1) \\ h(0)h(2) \\ h(1)h(2) \end{bmatrix} \quad (5)$$

Therefore, the monomial  $h(0)^2$  is now expressed.

#### Third step : expression of the third degree monomials

We now care about monomials  $h(0)^2h(1)$  and  $h(0)^2h(2)$ . These monomials can be expressed using the expression of the monomial  $h(0)^2$  in equation (5). In this equation, if we multiply  $h(0)^2$  by  $h(1)$ , monomials  $h(0)^2h(1)$  appears in the left hand side and monomials  $h(1)$ ,  $h(0)h(1)$ ,  $h(1)^2$ ,  $h(2)h(1)$ ,  $h(0)h(1)^2$ ,  $h(0)h(1)h(2)$  and  $h(1)^2h(2)$  appear in the right hand side. Among these monomials, one can distinguish those that are in the basis like  $h(1)$ ,  $h(0)h(1)$ ,  $h(2)h(1)$  and  $h(0)h(1)h(2)$ , those that are already expressed in the basis like  $h(1)^2$ , and those that are unknown like  $h(1)^2h(2)$  and  $h(0)h(1)^2$ . However, these unknown monomials are of the same kind as monomials  $h(0)^2h(1)$  and  $h(0)^2h(2)$ , and one can show that the expression of monomials  $h(0)^2h(1)$ ,  $h(0)^2h(2)$ ,  $h(1)^2h(0)$ ,  $h(1)^2h(2)$ ,  $h(2)^2h(0)$  and  $h(2)^2h(1)$  using (5) leads to :

$$\begin{bmatrix} h(0)^2h(1) \\ h(0)^2h(2) \\ h(1)^2h(0) \\ h(1)^2h(2) \\ h(2)^2h(0) \\ h(2)^2h(1) \end{bmatrix} = C \begin{bmatrix} 1 \\ h(0) \\ h(1) \\ h(2) \\ h(0)h(1) \\ h(0)h(2) \\ h(1)h(2) \\ h(0)h(1)h(2) \end{bmatrix} \quad (6)$$

#### Fourth step : expression of the fourth degree monomials

Using the same method as before, one can express the monomials like  $h(0)^2h(1)h(2)$  using equation (6).

We now have expressed all the monomials, and the multiplication matrix can then be built.

### 3.5. Choosing the channel estimate

Once the multiplication matrix is computed, the possible solutions are given by its eigenvectors. Then, the last step consists of choosing the solution that best matches the true channel.

The method used in this paper consists of comparing the circular-statistics of the observation with that given by each estimate. The solution that best matches is selected.

### 3.6. Identifiability

**Lemma 3.2** Suppose we look for a FIR channel  $H$  of length  $M$  from given second-order circular statistics, then the number of solutions is infinite because of a scalar phase indeterminacy. If the phase indeterminacy is fixed, the number of solutions is finite and equal to  $2^{M-1}$  if  $H$  is causal and equal to  $2^{2M-2}$  if  $H$  is not necessarily causal.

**Theorem 3.3** Suppose we look for a FIR channel  $H$  of length  $M$  from given second-order circular and non-circular statistics, then the number of solutions is finite and equal to :

- 2 if  $H$  has no real root,
- $2^{Q+1}$  if  $H$  has  $Q$  real roots and is causal,
- $2^{2Q+1}$  if  $H$  has  $Q$  real roots and is not necessarily causal.

When the source is MSK, the channel can be identified up to a sign.

*Proof.* The  $z$ -transform of the circular covariance  $c(n)$  of the output  $y(n)$  is equal to  $C(z) = H(z)H^*(1/z^*)$ . This shows that if  $H(z)$  is causal, it can be determined up to 2 indeterminacies. First,  $H(z)$  can only be determined up to a multiplicative constant phase factor. Second, if  $H(z)$  is transformed into  $H(z)\Phi(z)$  where  $\Phi(z)$  verifies  $\Phi(z)\Phi^*(1/z^*) = 1$ ,  $C(z)$  remains the same. It is well known that  $\Phi(z)$  is an all pass filter, i.e.  $\Phi(z)$  is of the form :

$$\Phi(z) = \prod_{i=1}^Q \frac{1 - a_i^* z^{-1}}{a_i - z^{-1}}$$

Since  $H(z)$  must be FIR and  $\Phi(z)$  is not FIR,  $H(z)\Phi(z)$  is FIR only if each pole of  $\Phi(z)$  is associated with a root of  $H(z)$ . As a consequence, there is a finite number of all-pass filters such that  $H(z)\Phi(z)$  is FIR. Therefore, if

the phase indeterminacy is fixed, there are  $2^{M-1}$  possible FIR filters that correspond to  $C(z)$ . This gives the first part of lemma 3.2.

If  $H(z)$  is non-causal, a third indeterminacy exists.  $C(z)$  has  $M - 1$  pairs of roots  $(b_i, 1/b_i^*)$  (if  $b_i$  is a root then  $1/b_i^*$  is also a root). Thus any  $H(z)$  built with  $M - 1$  roots, where each root  $h_j$  is equal to  $b_j$  or  $1/b_j^*$ , gives the same  $C(z)$ . The number of these channels is equal to  $2^{M-1}$ . Thus, when  $H(z)$  is non causal, the number of solutions is equal to  $2^{2M-2}$ . This gives the second part of lemma 3.2.

Suppose now that we also use the non-circular covariance  $\bar{c}(n)$  of  $y(n)$ . Its  $z$ -transform is equal to  $\bar{C}(z) = H(z)H(1/z)$ , if the input is white. This new constraint shows that the phase indeterminacy is reduced to a sign indeterminacy, and that the all-pass filter  $\Phi(z)$  must have real poles. As a consequence, if  $H(z)$  has no real roots, the all-pass filter  $\Phi(z)$  must be equal to  $\pm 1$ . In this case, there are only 2 solutions. If  $H(z)$  has  $Q$  real roots, one can use the results of lemma 3.2. The number of solutions is then equal to  $2^{Q+1}$  if  $H$  is causal, because the solutions are given up to a sign; if  $H$  is not necessarily causal the number of solutions is equal to  $2^{2Q+1}$ .

When the input source is MSK,  $\bar{C}(z) = H(z)H(-1/z)$ , which no all-pass filter but  $\Phi(z) = \pm 1$  satisfies. Therefore, the channel can be identified up to a sign if the input source is MSK.  $\square$

**Corollary 3.4** Suppose the circular and non-circular moments of the output  $y(n)$  are known and that the channel  $H(z)$  has no real roots. Then, there exists  $2^{M-1} C_{2M-2}^{M-1}$  possible solutions, each of them known up to a constant phase indeterminacy. They can be computed directly from the covariance  $C(z)$ . For each solution, the phase indeterminacy can be fixed with the non-circular moments, and the channel estimate is the channel that best matches the non-circular moments. This Corollary gives a new identification method.

## 4. COMPUTER RESULTS

The first tests have been run on a random FIR channel ( $M = 5$ ). At each run the channel is a realization of a Clarke filter in the typical urban mode and is excited by a MSK input. The performances are presented as a function of the SNR and of the length  $N$  of the observation block, and averaged over 500 runs.

Figure 1 shows the average Bit Error Rate obtained at the output of a Viterbi algorithm that uses our channel estimate. The solid, dashed and dashdotted lines correspond to block lengths  $N = 200$ ,  $N = 500$  and

$N = 1000$  respectively. These performances are compared to the average BER obtained with the true channel (dotted line).

For high SNRs and  $N = 200$  or  $N = 500$  the results show the effects of statistic estimation errors. These effects disappear for  $N = 1000$ , where the performances exhibit a loss of  $2dB$  compared to the true channel results.

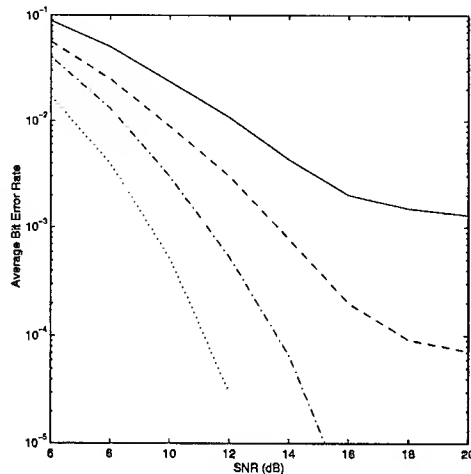


Figure 1: Average Bit Error Rate at the output of a Viterbi algorithm using our channel estimate.

A second test has been run on a random FIR channel of length  $M = 3$ . Figure 2 shows the average Bit Error Rate obtained at the output of a Viterbi algorithm that uses our channel estimate when  $\hat{M} = 3$  (solid line) and  $\hat{M} = 5$  (dashdotted line). These two performances are compared to the one obtained when the true channel is used (dotted line). Hence, this test illustrates the loss of performance encountered in presence of over-determination.

## 5. CONCLUDING REMARKS

In this paper, we presented a new blind identification method based on the non-circular moments of the observation. These moments yield a polynomial system that can be solved by computing the eigenvectors of a multiplication matrix. Identifiability results are presented when a FIR channel is searched for, from both circular and non-circular moments. Finally, computer results show that the behaviour of our algorithm depends on the estimation of the moments.

## 6. REFERENCES

[1] P. COMON, O. GRELLIER, B. MOURRAIN, "Closed-form blind channel identification with

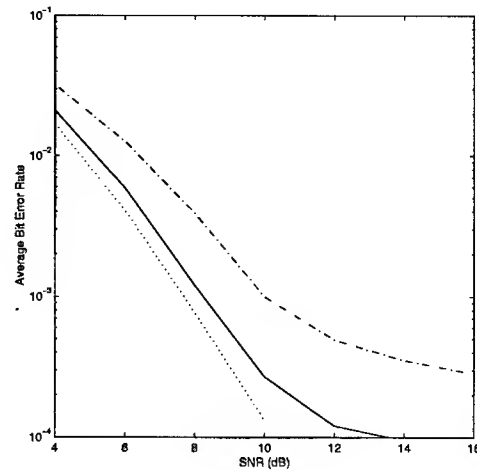


Figure 2: Average Bit Error Rate at the output of a Viterbi algorithm using our channel estimate in presence of over-determination

MSK inputs", in *Asilomar Conference*, Pacific Grove, California, November 1-4 1998, pp. 1569-1573, Invited session.

- [2] J. HARRIS, *Algebraic Geometry, a first course*, vol. 133 of *Graduate Texts in Math.*, Springer, 1992.
- [3] T.-H. LI, "Blind identification and deconvolution of linear systems driven by binary random sequences", *Trans. on Inf. Theory*, vol. 38, no. 1, pp. 26-38, Jan 1992.
- [4] Y. LI, Z. Ding, "Arma system identification based on second-order cyclostationary", *IEEE Trans. Sig. Proc.*, vol. 42, no. 12, pp. 3483-3494, Dec. 1994.
- [5] O. SHALVI, E. WEINSTEIN, "New criteria for blind deconvolution of nonminimum phase systems", *IEEE Trans. Inf. Theory*, vol. 36, no. 2, pp. 312-321, Mar. 1990.
- [6] L. TONG, G. XU, T. KAILATH, "Blind identification and equalization based on second-order statistics: a time domain approach", *IEEE Trans. Inf. Theory*, vol. 40, no. 2, pp. 340-349, Mar. 1994.
- [7] J. TUGNAIT, "Comments on 'New criteria for blind deconvolution of nonminimum phase systems'", *IEEE Trans. Inf. Theory*, vol. 38, no. 1, pp. 210-213, Jan. 1992.
- [8] D. YELLIN, B. PORAT, "Blind identification of FIR systems excited by discrete-alphabet inputs", *IEEE Trans. Sig. Proc.*, vol. 41, no. 3, pp. 1331-1339, 1993.

# THE ROLE OF SECOND-ORDER STATISTICS IN BLIND EQUALIZATION OF NONLINEAR CHANNELS

Roberto López-Valcarce and Soura Dasgupta

Dept. Electrical and Computer Engineering  
University of Iowa, 52242-1595 IA, USA  
valcarce@icaen.uiowa.edu, dasgupta@eng.uiowa.edu

## ABSTRACT

We explore the utility of second-order statistics for blind equalization of nonlinear channels. Although any SOS-based method can only identify the channel to within a mixing matrix (at best), sufficient conditions are given to ensure that the ambiguity is at a level that still allows equalization. These conditions are satisfied by a wider class of inputs than those satisfying the conditions derived in previous works.

## 1. INTRODUCTION

In recent years blind equalization of single-input multiple-output (SIMO) linear channels has received considerable attention, motivated by the fact that these channels can be perfectly equalized if the subchannels are coprime and the equalizer is long enough, and that the equalizer can be obtained from the second-order statistics (SOS) of the received signal [6].

With a few exceptions [1, 5, 8], almost all the available literature on blind equalization is devoted to the linear channel case. However, many real world communication systems, such as digital satellite and radio links, high-density magnetic and optical storage channels, etc., exhibit a considerable degree of nonlinearity. Hence it is of interest to address the issue of blind equalization of *nonlinear channels*. The SIMO channel model that we consider here has the following form:

$$x_n = \sum_{j=0}^{L_0} h_{0j} a_{n-j} + \sum_{i=1}^D \sum_{j=0}^{L_i} h_{ij} z_n^{(i)} + \eta_n, \quad (1)$$

where  $\{a_n\}$  is the scalar, stationary input, the terms  $z_n^{(i)} = f_i(a_n, a_{n-1}, \dots)$  are known scalar-valued nonlinear causal functions of  $\{a_n\}$ ,  $h_{ij}$  are  $K \times 1$  coefficient vectors, and  $\eta_n$ ,  $x_n$  are  $K \times 1$  signal vectors representing an additive disturbance and the observed signal, respectively; the number of subchannels is  $K$ . The noise

$\{\eta_n\}$  and the signal  $\{a_n\}$  are assumed to be independent. This model accommodates, for example, polynomial approximations of nonlinear channels (Volterra models), though the 'basis functions'  $\{z_n^{(i)}\}$  need not be monomials in principle.

We seek conditions under which zero-forcing (ZF) linear equalizers for the class of channels (1) can be designed using only the SOS of  $\{x_n\}$ . SOS-based methods are often preferred since they can perform their tasks with relatively short data records. The fact that linear finite impulse response (FIR) systems can perform ZF equalization of nonlinear SIMO Volterra channels under certain conditions was first pointed out in [1], together with a blind, deterministic approach for equalizer design. Although the method is simple, it has been shown in [3] that the conditions in [1] are in fact conservative. In [3] we have given certain conditions on the input statistics that suffice to determine a linear equalizer for (1). In this paper we significantly expand the results of [3] by giving new conditions that accommodate a wider class of inputs.

Another SOS-based approach is suggested in [5], inspired in the method of [9] for linear channels. However, this method requires that every nonlinear subchannel be linearizable by an FIR Volterra system of known order and memory, which is in general not possible, especially if each subchannel is modeled as an FIR Volterra system itself (a common practice).

In principle the model (1) could be seen as a linear multiple-input multiple-output (MIMO) system by treating the  $\{z_n^{(i)}\}_{i=1}^D$  as additional inputs. Although SOS-based techniques exist for equalization within such a framework [6], these techniques usually assume that the different inputs are uncorrelated (which is no longer true in our setting), and they only resolve the inputs to within a mixing matrix. In the current context, as  $z_n^{(i)}$  are functions of  $\{a_n\}$ , this would mean that only a memoryless nonlinear function of the input could be obtained. The results of this work show that under right

conditions the structure of the mixing matrix permits obtaining linear ZF equalizers. These conditions are on the statistical properties of the symbols  $\{a_n\}$  and the remaining basis functions  $\{z_n^{(i)}\}$ . Therefore they can be checked *a priori* in order to determine whether a given channel structure can be equalized from SOS.

## 2. PROBLEM STATEMENT

We denote transpose and conjugate transpose by  $(\cdot)^T$ ,  $(\cdot)^H$  respectively. By collecting  $N$  successive observations into  $X_n^T = [x_n^T \cdots x_{n-N+1}^T]$ , one can write

$$X_n = \mathcal{F}S_n + V_n, \quad (2)$$

where  $V_n^T = [\eta_n^T \cdots \eta_{n-N+1}^T]$  is the noise vector, the signal regressor  $S_n^T = [(S_n^{(1)})^T \cdots (S_n^{(D)})^T]$  with

$$S_n^{(1)} = [a_n \ a_{n-1} \ \cdots \ a_{n-L_0-N+1}]^T, \quad (3)$$

$$S_n^{(2)} = [z_n^{(1)} \ \cdots \ z_{n-L_1-N+1}^{(1)} \ | \ \cdots \ | \ z_n^{(D)} \ \cdots \ z_{n-L_D-N+1}^{(D)}]^T, \quad (4)$$

and the channel matrix  $\mathcal{F} = [\mathcal{F}_0 \ \mathcal{F}_1 \ \cdots \ \mathcal{F}_D]$ , with every  $\mathcal{F}_i$  block Toeplitz,

$$\mathcal{F}_i = \begin{bmatrix} h_{i0} & \cdots & h_{iL_i} & & \\ & \ddots & & \ddots & \\ & & h_{i0} & \cdots & h_{iL_i} \end{bmatrix} \quad KN \times (N+L_i).$$

For convenience, let  $k_1 = N + L_0$ , which is the size of  $S_n^{(1)}$ , the linear part of the regressor; and  $k_2 = L_1 + \cdots + L_D + DN$ , which is the size of  $S_n^{(2)}$  (thus  $S_n$  is  $(k_1 + k_2) \times 1$ ).

From (2), one has

$$C_x(k) = \text{cov}(X_n, X_{n-k}) = \mathcal{F}C_s(k)\mathcal{F}^H + C_v(k), \quad (5)$$

with  $C_s(k) = \text{cov}(S_n, S_{n-k})$ ,  $C_v(k) = \text{cov}(V_n, V_{n-k})$  the signal and noise covariance matrices. We adopt the following standard assumptions:

**A1:** The channel matrix  $\mathcal{F}$  has full column rank.

**A2:**  $\{\eta_n\}$  is zero-mean, white, with covariance  $\sigma_\eta^2 I_K$ .

**A3:** The covariance matrix  $C_s(0)$  is positive definite.

A necessary condition for **A1** to hold is  $K > D + 1$ , which parallels the 'more outputs than inputs' condition in blind identification of MIMO channels. Observe that **A1** ensures the existence of vectors  $g_\delta$  such that  $g_\delta^H \mathcal{F} = e_\delta^H$ , where  $e_\delta$  is the  $\delta$ -th unit vector (counting from zero). Thus in the noiseless case, for  $0 \leq \delta \leq k_1 - 1$  one has  $g_\delta^H X_n = a_{n-\delta}$  so that these vectors provide ZF linear equalizers. From these, minimum mean-square-error equalizers can be obtained [4].

Under **A2**,  $\sigma_\eta^2$  can be estimated as the smallest eigenvalue of  $C_x(0)$ . Thus the effect of the noise can be removed from  $C_x(k)$ . Henceforth we shall assume that  $C_x(k) = \mathcal{F}C_s(k)\mathcal{F}^H$ . The problem under consideration can be posed as follows:

**Blind Equalizability Problem:** Let  $\tilde{\mathcal{F}}$  be a matrix of the same size as  $\mathcal{F}$  such that

$$\tilde{\mathcal{F}}C_s(k)\tilde{\mathcal{F}}^H = \mathcal{F}C_s(k)\mathcal{F}^H, \quad k = 0, 1, \dots, \bar{k}. \quad (6)$$

We say that  $\tilde{\mathcal{F}}$  is compatible with the second order statistics of  $X_n$ . Determine conditions under which a ZF equalizer  $g_\delta$  for any compatible  $\tilde{\mathcal{F}}$  is also a ZF equalizer for  $\mathcal{F}$ . That is,

$$g_\delta^H \tilde{\mathcal{F}} = e_\delta^H \implies g_\delta^H \mathcal{F} = c e_\delta^H, \quad (7)$$

with  $0 \leq \delta \leq k_1 - 1$  and  $c \neq 0$ .

This was solved in [7] for the particular case of linear channels with white inputs, for which if  $\tilde{\mathcal{F}}$  is compatible with  $\bar{k} = 1$ , then  $\tilde{\mathcal{F}} = e^{j\theta} \mathcal{F}$  so that (7) holds. It is our goal to extend this result to the class of channels (1) and colored inputs.

## 3. THE AMBIGUITY MATRIX

Observe that assumption **A3** allows us to write  $C_s(0) = QQ^H$  where  $Q$  is nonsingular (not necessarily unique). Introduce the *normalized* channel and signal covariance matrices respectively as

$$F = \mathcal{F}Q, \quad \bar{C}_s(k) = Q^{-1}C_s(k)Q^{-H}. \quad (8)$$

Using (8), the covariance matrices  $C_x(k)$  become

$$C_x(k) = F\bar{C}_s(k)F^H, \quad \text{with } \bar{C}_s(0) = I. \quad (9)$$

Similarly, if  $\tilde{\mathcal{F}}$  is compatible, let  $\tilde{F} = \tilde{\mathcal{F}}Q$ , so that  $\tilde{F}$  satisfies

$$\tilde{F}\bar{C}_s(k)\tilde{F}^H = F\bar{C}_s(k)F^H, \quad 0 \leq k \leq \bar{k}. \quad (10)$$

For  $k = 0$ , (10) reads as  $\tilde{F}\tilde{F}^H = FF^H$ . Since  $F$  has full column rank, this implies  $\tilde{F} = FP$  for some unitary matrix  $P$ . Thus the corresponding (unnormalized) compatible channel matrix must satisfy

$$\tilde{\mathcal{F}} = \tilde{F}Q^{-1} = \mathcal{F}(QPQ^{-1}), \quad (11)$$

which shows that any compatible channel matrix is related to the true channel via a mixing matrix of the form  $\tilde{P} = QPQ^{-1}$ . Observe that although  $P$  is unitary, in general  $\tilde{P}$  is not. Let us introduce the concept of admissibility.



**Definition 1 (Admissibility)** A  $(k_1 + k_2)$ -square matrix  $T$  is said to be admissible if it is of the form

$$T = \begin{bmatrix} \Lambda & 0 \\ * & * \end{bmatrix}, \quad \Lambda \text{ } k_1 \times k_1 \text{ diagonal invertible.} \quad (12)$$

with the asterisks indicating irrelevant values. Note that if  $T$  is admissible and invertible, so is  $T^{-1}$ ; and any function of an admissible matrix is admissible.

Observe that if  $\tilde{\mathcal{F}} = \mathcal{F}\tilde{P}$  is compatible with  $\tilde{P} = QPQ^{-1}$  admissible, then the condition (7) is satisfied. Thus resolution of the channel matrix to within this ambiguity suffices for equalization purposes. We now ask, when is this resolution possible?

To answer this question we must explore the constraints that the conditions (10) impose on the matrix  $P$ . Substituting  $\mathcal{F} = FP$  into (10) and using the fact that  $F$  has full rank, these constraints can be written as

$$P\tilde{C}_s(k) = \tilde{C}_s(k)P, \quad 1 \leq k \leq \bar{k}. \quad (13)$$

That is,  $P$  must commute with the normalized source covariance matrices  $\tilde{C}_s(1), \dots, \tilde{C}_s(\bar{k})$ .

#### 4. A SIMPLIFIED EQUALIZABILITY TEST

Determining the general form of all unitary matrices  $P$  that satisfy (13) requires solving a linear set of equations with quadratic constraints. Fortunately, this problem can be replaced by one of solving a linear set of equations with linear constraints. First recall that any unitary matrix  $P$  can be written as  $P = e^{jW}$  where  $W$  is a Hermitian matrix with eigenvalues in  $[0, 2\pi)$  [2]. Secondly, we have the following result [3]:

**Theorem 1** Let  $W$  be  $(k_1 + k_2)$ -square Hermitian and  $P = e^{jW}$ . Then  $P$  and  $\tilde{C}_s(k)$  commute if and only if  $W$  and  $\tilde{C}_s(k)$  commute.

Hence the problem can be broken into these three steps:

1. Select a square root  $Q$  of  $C_s(0)$ .
2. Find all Hermitian matrices  $W$  commuting with  $\tilde{C}_s(k) = Q^{-1}C_s(k)Q^{-H}$  for  $1 \leq k \leq \bar{k}$ .
3. Check whether for these matrices  $W$ ,  $QWQ^{-1}$  is admissible. If so, the channel can be equalized using second-order statistics.

The usefulness of theorem 1 is revealed in that steps 2 and 3 above are much easier to solve for Hermitian matrices than for unitary matrices.

As noted above, the matrix  $Q$  such that  $C_s(0) = QQ^H$  is not unique. Although it is true that an adequate choice of  $Q$  can considerably simplify the test for

SOS-based equalizability, as discussed in section 5, it must be pointed out that the result of the test is independent of  $Q$ . This is because all square roots can be parameterized as  $Q = Q_0U$ , where  $Q_0$  is a particular solution and  $U$  is any unitary matrix. Consequently, a unitary  $P_0$  commutes with  $Q_0^{-1}C_s(k)Q_0^{-H}$  if and only if  $P = U^H P_0 U$ , which is unitary, commutes with  $Q^{-1}C_s(k)Q^{-H}$ . In addition, one has

$$QPQ^{-1} = Q_0 P_0 Q_0^{-1},$$

so that admissibility of  $QPQ^{-1}$  is equivalent to that of  $Q_0 P_0 Q_0^{-1}$ . Thus equalizability does not depend on the specific square root  $Q$ .

Our goal now is to determine sufficient conditions in order to ensure success of the SOS-based equalizability test *a priori*.

#### 5. MAIN RESULTS

It will be especially useful to consider square roots  $Q$  which are block lower triangular (with block partition corresponding to linear and nonlinear parts of  $S_n$ , as in (12)), for the following reason: suppose that the Hermitian matrices  $W$  solving step 2 of the equalizability test are block diagonal. Then  $P = e^{jW}$  are block diagonal, and thus if  $Q$  was block lower triangular, the mixing matrices  $\tilde{P} = QPQ^{-1}$  will be block lower triangular as well. Having  $\tilde{P}$  block lower triangular (i.e. of the form (12) but with  $\Lambda$  not necessarily diagonal) is the first step towards admissibility: its significance is that a linear ZF equalizer  $g_d$  for the compatible matrix  $\tilde{\mathcal{F}}$  ( $g_d^H \tilde{\mathcal{F}} = e_d^H$ ,  $0 \leq d \leq k_1 - 1$ ), although not a ZF equalizer for the true channel  $\mathcal{F}$ , still removes all the nonlinear ISI if  $\tilde{P}$  is block lower triangular since  $g_d^H \mathcal{F} = e_d^H \tilde{P}^{-1}$ .

Before presenting block lower triangular options for the square root  $Q$ , we give the following result which ensures the block diagonal property of the Hermitian matrices  $W$  that commute with  $C_s(1)$ .

**Theorem 2** Assume that there exists a matrix  $Q$  such that  $C_s(0) = QQ^H$  and

$$\tilde{C}_s(1) = Q^{-1}C_s(1)Q^{-H} = \begin{bmatrix} C_{11} & 0 \\ C_{21} & C_{22} \end{bmatrix}, \quad (14)$$

with  $C_{ij}$  having size  $k_i \times k_j$ . Suppose that either (i)  $C_{11}, C_{22}$  do not share any eigenvalues; or (ii)  $C_{21} = 0$ , and  $C_{11}, C_{22}$  do not share any elementary Jordan block in their Jordan decompositions. Then any Hermitian  $W$  commuting with  $\tilde{C}_s(1)$  must be of the form

$$W = \begin{bmatrix} W_{11} & 0 \\ 0 & W_{22} \end{bmatrix}, \quad (15)$$

with  $W_{ii}$  Hermitian of size  $k_i \times k_i$ .

With this result in mind, we shall focus on block triangular square roots  $Q$  and look for conditions under which (14) is satisfied. Let  $A_{ij} = \text{cov}(S_n^{(i)}, S_n^{(j)})$  (which has size  $k_i \times k_j$ ) so that

$$C_s(0) = \begin{bmatrix} A_{11} & A_{12} \\ A_{12}^H & A_{22} \end{bmatrix}.$$

Define the Schur complement  $A_0 = A_{22} - A_{12}^H A_{11}^{-1} A_{12}$ , which is positive definite. The following choice of  $Q$  will prove particularly useful:

$$Q = \begin{bmatrix} A_{11}^{1/2} & 0 \\ A_{12}^H A_{11}^{-H/2} & A_0^{1/2} \end{bmatrix}, \quad (16)$$

where  $A_{11}^{1/2}$ ,  $A_0^{1/2}$  are square roots of  $A_{11}$  and  $A_0$  respectively, i.e.  $A_{11} = A_{11}^{1/2} A_{11}^{H/2}$ ,  $A_0 = A_0^{1/2} A_0^{H/2}$ . The following results give sufficient conditions on the source statistics and the channel nonlinearities in order to have  $\tilde{P}$  admissible.

**Theorem 3** Suppose that the symbol sequence  $\{a_n\}$  is an autoregressive (AR) process of order not exceeding  $k_1$  with independent, identically distributed (iid) innovations, i.e. it is generated by means of all-pole filtering of an iid process  $\{w_n\}$  as follows:

$$a_n = w_n - \sum_{i=1}^{k_1} \gamma_i a_{n-i}. \quad (17)$$

Then with  $Q$  as in (16), the corresponding matrix  $\tilde{C}_s(1)$  is block triangular as in (14).

In addition, suppose that the diagonal blocks of  $\tilde{C}_s(1)$  do not share any eigenvalue. Then for all Hermitian matrices  $W$  commuting with  $\tilde{C}_s(1)$ ,  $QWQ^{-1}$  is admissible.

This result can be understood as follows. The autoregressive condition on the symbols  $\{a_n\}$  provides the desired block triangular structure (14) for  $\tilde{C}_s(1)$ . If the diagonal blocks of  $\tilde{C}_s(1)$  do not share any eigenvalue, then one can conclude from theorem 2 that the Hermitian matrices  $W$  must be block diagonal. Theorem 3 tells us that in addition to this, these  $W$  are such that  $\tilde{P} = Qe^{jW}Q^{-1}$  is admissible. Thus for AR symbols, it suffices for equalizability to check the eigenvalue condition on  $\tilde{C}_s(1)$ . Observe that iid symbol sequences constitute a particular class of AR processes for which  $\gamma_i = 0$  in (17).

The next result provides similar conclusions under different conditions:

**Theorem 4** Suppose that the symbols  $\{a_n\}$  are Gaussian, and that the memory of the nonlinear part of the

channel does not exceed that of the linear part. Then with  $Q$  as in (16), the corresponding matrix  $\tilde{C}_s(1)$  is block diagonal, i.e. as in (14) with  $C_{21} = 0$ .

In addition, suppose that the diagonal blocks of  $\tilde{C}_s(1)$  do not share any elementary Jordan block in their Jordan decompositions. Then for all Hermitian matrices  $W$  commuting with  $\tilde{C}_s(1)$ ,  $QWQ^{-1}$  is admissible.

We must remark that having the memory of the nonlinear part no larger than that of the linear part is not the same as saying that  $L_0 \geq L_i$ ,  $i = 1, \dots, D$  in (1). (This is because the terms  $z_n^{(i)}$  need not be memoryless). Instead, this memory requirement can be interpreted as having

$$S_n^{(2)} = f(S_n^{(1)}) \quad \text{with } f(\cdot) \text{ a memoryless mapping.}$$

According to theorem 4, the coloring of the symbols need not be restricted to AR models if the symbols are Gaussian. Gaussianity also result in a less stringent eigenvalue condition on  $\tilde{C}_s(1)$  for admissibility (since its diagonal blocks may now share eigenvalues as long as their respective Jordan blocks have different sizes). On the other hand, the conditions on theorem 4 include the memory limitation on the nonlinear part of the channel, while no such constraint was present in theorem 3 for AR symbols.

Let us define  $B_{ij} = \text{cov}(S_n^{(i)}, S_{n-1}^{(j)})$ ,  $i, j = 1, 2$ , and

$$B_0 = B_{22} - A_{12}^H A_{11}^{-1} B_{11} A_{11}^{-1} A_{12}. \quad (18)$$

It can be shown that under the conditions of theorem 4, the diagonal blocks  $C_{11}$  and  $C_{22}$  of  $\tilde{C}_s(1)$  in (14) are respectively similar to  $B_{11}A_{11}^{-1}$  and  $B_0A_0^{-1}$ . Thus one must compare the elementary Jordan blocks in the Jordan decompositions of these matrices. Observe that  $B_{11}A_{11}^{-1}$  is a companion matrix associated to the forward prediction error filter of order  $k_1$  for the process  $\{a_n\}$  (If the symbol sequence is white, then  $B_{11}A_{11}^{-1}$  reduces to the shift matrix with ones in the first sub-diagonal and zeros elsewhere, as in [3]). Therefore all eigenvalues lie inside the unit circle, and there is only one elementary Jordan block associated to each distinct eigenvalue [2]. On the other hand, the Jordan structure of  $B_0A_0^{-1}$  appears to be more intricate due to possible interactions between different terms  $z_n^{(i)}$ .

Finally, when the symbol sequence  $\{a_n\}$  and the nonlinear terms are uncorrelated, the Jordan block condition is sufficient for equalizability, regardless of the color or distribution of the symbols:

**Theorem 5** Suppose that  $\text{cov}(a_n, z_{n-k}^{(i)}) = 0$  for all  $k$  and for  $1 \leq i \leq D$ . Then with  $Q$  as in (16), the matrix  $\tilde{C}_s(1)$  is block diagonal:  $\tilde{C}_s(1) = \begin{bmatrix} C_{11} & 0 \\ 0 & C_{22} \end{bmatrix}$ , with

$C_{ii}$  of size  $k_i \times k_i$ . If  $C_{11}$  and  $C_{22}$  do not share any elementary Jordan block in their Jordan decompositions, then for any Hermitian  $W$  commuting with  $\tilde{C}_s(1)$ , the matrix  $QWQ^{-1}$  is admissible.

## 6. EXAMPLES

Suppose that the symbols  $\{a_n\}$  are generated by means of  $a_n = w_n + w_{n-1}$ , where  $\{w_n\}$  is a real, zero-mean, iid process, symmetrically distributed around the origin. Let  $D = 1$  and  $z_n^{(1)} = z_n = a_n^2$ . Then one has  $\text{cov}(a_n, z_{n-k}) = 0$  for all  $k$ , so that we are in the conditions of theorem 5. The matrices  $C_{11}$  and  $C_{22}$  for this case are similar to the companion matrices associated to the forward prediction error filters of orders  $k_1$  and  $k_2$  for the processes  $\{a_n\}$  and  $\{z_n\}$ , respectively. In view of theorem 5, if the transfer functions of these prediction error filters are coprime, then SOS-based equalizability is ensured. Both  $\{a_n\}$  and  $\{z_n\}$  are first-order Moving Average processes with autocovariance coefficients

$$\begin{aligned}\rho_1 &= \frac{\text{cov}(a_n, a_{n-1})}{\text{cov}(a_n, a_n)} = \frac{1}{2}, \\ \rho_2 &= \frac{\text{cov}(z_n, z_{n-1})}{\text{cov}(z_n, z_n)} = \frac{1 - \alpha}{2(1 + \alpha)},\end{aligned}$$

where  $\alpha = E^2[w_n^2]/E[w_n^4]$ . For example, for Gaussian  $\{w_n\}$ ,  $\alpha = \frac{1}{3}$  and  $\rho_2 = \frac{1}{4}$ ; and for equiprobable  $w_n = \pm 1$  (a BPSK signal), one has  $\alpha = 1$  and  $\rho_2 = 0$ , so that  $\{z_n\}$  is white. Moreover, note that since  $\rho_1 = \frac{1}{2}$ ,  $\lambda = 0$  is never an eigenvalue of  $C_{11}$ . Thus we conclude that for BPSK  $\{w_n\}$  this channel is always SOS-equalizable, irrespective of the kernel memories  $L_0$  and  $L_1$ .

Now suppose that  $\{w_n\}$  is Gaussian and that  $z_n = a_n^3$ . The processes  $\{a_n\}$  and  $\{z_n\}$  are not uncorrelated any more; hence theorem 5 does not apply. However, if  $L_1 \leq L_0$ , the conditions of theorem 4 are satisfied. Suppose  $L_1 = L_0$ . The matrices  $C_{11}$  and  $C_{22}$  are now similar to the companion matrices associated to the forward prediction error filters of order  $k_1 = k_2 = N + L_0$  for two MA(1) processes with autocovariance coefficients  $\rho_1 = \frac{1}{2}$  and  $\rho_2 = -\frac{7}{16}$ . The transfer functions of these filters are always coprime, so that the channel is SOS-equalizable.

## 7. CONCLUSIONS

The problem of blindly equalizing a SIMO nonlinear FIR channel using second-order statistics of the observed signal has been considered. We have presented sufficient conditions on the statistics of the symbol sequence and the channel structure (which can be checked *a priori*) allowing the design of linear FIR zero-forcing

equalizers. These conditions considerably expand previous results by accommodating a wider class of inputs.

Our equalizability conditions do not describe how the linear equalizers can be found from the output SOS. Algorithm development exploiting the results presented is the next logical step and is currently under study.

## REFERENCES

- [1] G. B. Giannakis and E. Serpedin, "Linear multi-channel blind equalizers of nonlinear FIR Volterra channels", *IEEE Trans. Signal Proc.*, vol. 45 no. 1, pp. 67-81, Jan. 1997.
- [2] P. Lancaster and M. Tismenetsky, *The theory of matrices*, Academic Press, 1985.
- [3] R. López-Valcarce and S. Dasgupta, "Blind identifiability/equalizability of single input multiple output nonlinear channels from second order statistics", to be presented in 2000 IEEE ICASSP, Istanbul, Turkey.
- [4] C. Papadias and D. T. M. Slock, "Fractionally spaced equalization of linear polyphase channels and related blind techniques based on multichannel linear prediction", *IEEE Trans. Signal Proc.*, vol 47 no. 3, pp. 641-654, March 1999.
- [5] G. M. Raz and B. D. Van Veen, "Blind equalization and identification of nonlinear and IIR systems—A Least Squares approach", *IEEE Trans. Signal Processing*, vol. 48 no. 1, pp. 192-200, January 2000.
- [6] L. Tong and S. Perreau, "Multichannel blind identification: from subspace to maximum likelihood methods", *Proc. IEEE*, vol. 86 no. 10, pp. 1951-1968, Oct. 1998.
- [7] L. Tong, G. Xu and T. Kailath, "Blind identification and equalization based on second-order statistics: a time-domain approach", *IEEE Trans. Information Theory*, vol. 40, no. 2, pp. 340-350, March 1994.
- [8] M. Tsatsanis and H. Cirpan, "Blind identification of nonlinear channels excited by discrete alphabet inputs", *Proc. 1996 IEEE Signal Proc. Workshop Stat. Signal Array Proc.*, vol. 1, pp. 176-179, Corfu, Greece.
- [9] G. Xu, H. Liu, L. Tong and T. Kailath, "A Least-Squares approach to blind channel identification" *IEEE Trans. Signal Proc.*, vol. 43, no. 12, pp. 2982-2993, Dec. 1995.

# ON SUPER-EXPONENTIAL ALGORITHM, CONSTANT MODULUS ALGORITHM AND INVERSE FILTER CRITERIA FOR BLIND EQUALIZATION

*Chong-Yung Chi, Ching-Yung Chen and Bin-Way Li*

Department of Electrical Engineering  
National Tsing Hua University, Hsinchu, Taiwan, R.O.C.  
Tel: 886-3-5731156, Fax: 886-3-5751787, E-mail: cychi@ee.nthu.edu.tw

## ABSTRACT

Super-exponential algorithm (SEA), constant modulus algorithm (CMA) and inverse filter criteria (IFC) using higher-order statistics have been widely used for blind equalization. Chi, Feng and Chen have reported that SEA and IFC are equivalent under certain conditions. In this paper, we further prove that SEA, IFC and CMA are equivalent under certain conditions, and their convergence speed and computational load can be significantly improved as the given data are preprocessed by the well-known lattice linear prediction error (LPE) filter for both off-line processing and adaptive processing. Some simulation results are presented to support the analytic results and the proposed off-line and adaptive implementations.

## 1. INTRODUCTION

Blind equalization (deconvolution) is a signal processing procedure to recover the desired independent identically distributed (i.i.d.) non-Gaussian signal, denoted by  $u[n]$ , that is transmitted through an unknown linear time-invariant (LTI) channel, denoted by  $h[n]$ , with only measurements

$$\begin{aligned} x[n] &= u[n] * h[n] + w[n] \\ &= \sum_{k=-\infty}^{\infty} h[k]u[n-k] + w[n] \end{aligned} \quad (1)$$

where  $w[n]$  is additive noise. The problem of blind equalization arises comprehensively in a variety of applications such as digital communications, seismic deconvolution, speech modeling and synthesis, ultrasonic nondestructive evaluation and image restoration.

The FIR linear equalizer of order  $L$ , denoted by  $v[n]$ , has been widely used to process  $x[n]$  such that the

This work was supported by the National Science Council under Grant NSC-89-2213-E-007-073.

equalizer output

$$\begin{aligned} e[n] &= x[n] * v[n] = \sum_{k=0}^L v[k]x[n-k] \\ &= u[n] * g[n] + w[n] * v[n] \quad (\text{by (1)}) \end{aligned} \quad (2)$$

approximates  $\alpha u[n - \tau]$  ( $\alpha \neq 0$ ) where

$$g[n] = h[n] * v[n] \quad (3)$$

is the overall system after equalization. The amount of intersymbol interference (ISI) defined as [1]

$$\text{ISI}\{g(n)\} = \frac{\sum_n |g(n)|^2 - \max\{|g(n)|^2, \forall n\}}{\max\{|g(n)|^2, \forall n\}} \quad (4)$$

has been used as a performance index of the designed  $v[n]$ . The smaller ISI implies the better performance.

A number of blind equalization algorithms using higher-order statistics (cumulants and moments) have been reported for designing  $v[n]$  such as the well-known constant modulus algorithm (CMA) [2], inverse filter criteria (IFC) [3] and super-exponential algorithm (SEA) [1]. Chi, Feng and Chen [4] have reported the equivalence of IFC and SEA under certain conditions. In this paper, we further prove the equivalence of IFC, SEA and CMA under certain conditions, thus sharing some properties reported in [4-7] under these conditions. Furthermore, efficient implementations of these algorithms with preprocessing by linear prediction error (LPE) filter are presented including off-line processing and adaptive processing.

## 2. BACKGROUND

Let  $\text{cum}\{x_1, \dots, x_p\}$  denote the  $p$ th-order joint cumulant of random variables  $x_1, \dots, x_p$ , and  $\text{cum}\{e[n] : p, \dots\} = \text{cum}\{x_1 = e[n], \dots, x_p = e[n], \dots\}$ . For ease of later use, let us define the following notations

$$\begin{aligned}
\mathbf{v} &= (v[0], v[1], \dots, v[L])^T \\
\mathbf{x}[n] &= (x[n], x[n-1], \dots, x[n-L])^T \\
f_k[n] &: k\text{th-order forward prediction error} \\
b_k[n] &: k\text{th-order backward prediction error} \\
\mathbf{f}_k[n] &= (f_k[n], f_k[n-1], \dots, f_k[n-L])^T \\
\mathbf{b}[n] &= (b_0[n], b_1[n], \dots, b_L[n])^T \\
C_{p,q}^u &= \text{cum}\{u[n] : p, u^*[n] : q\} \\
\text{sgn}(\alpha) &: \text{sign of real-valued } \alpha
\end{aligned}$$

### 2.1. Lattice LPE Filter

The  $k$ th-order lattice LPE filter with reflection coefficients  $\rho_1, \rho_2, \dots, \rho_k$ , simultaneously provides the forward prediction error  $f_k[n]$  and backward prediction error  $b_k[n]$ , that can be expressed as follows:

$$f_k[n] = \sum_{i=0}^k a_k[i]x[n-i] \quad (5)$$

$$b_k[n] = \sum_{i=0}^k a_k^*[k-i]x[n-i] \quad (6)$$

where the superscript  $*$  denotes complex conjugation,  $a_k[0] = 1$  and  $a_k[1], a_k[2], \dots, a_k[k]$ , can be obtained from  $\rho_1, \rho_2, \dots, \rho_k$  through the computationally efficient Levinson-Durbin recursion. Two facts regarding  $f_k[n]$  and  $b_k[n]$  are as follows [8]:

(F1) The  $k$ th-order LPE filter  $a_k[i]$  is a whitening filter as  $k$  is sufficiently large, i.e.,

$$\mathbf{R}_{f_k} = E[\mathbf{f}_k[n]\mathbf{f}_k^H[n]] \cong \sigma_{f,k}^2 \mathbf{I} \quad (7)$$

for sufficiently large  $k$  where  $\mathbf{I}$  is the  $(L+1) \times (L+1)$  identity matrix.

(F2)  $\mathbf{x}[n]$  and  $\mathbf{b}[n]$  are causally invertible and

$$\mathbf{R}_b = E[\mathbf{b}[n]\mathbf{b}^H[n]] = \text{diag}(P_0, P_1, \dots, P_L) \quad (8)$$

### 2.2. CMA

The CMA [2] finds the optimal equalizer  $v[n]$  by minimizing the following cost function

$$J_{\text{CM}}(\mathbf{v}) = E[(\gamma - |e[n]|^2)^2] \quad (9)$$

where  $\gamma = E[|u[n]|^4]/E[|u[n]|^2]^2$ . However, one has to resort to iterative optimization algorithms for searching the optimum  $\mathbf{v}$ .

### 2.3. SEA

Shalvi and Weinstein's SEA( $p, q$ ) [1] is an iterative algorithm that updates  $\mathbf{v}$  by the following equations at each iteration:

$$\mathbf{v} = \mathbf{R}_x^{-1} \mathbf{d} / \|\mathbf{R}_x^{-1} \mathbf{d}\| \quad (10)$$

where  $\mathbf{R}_x = E[\mathbf{x}[n]\mathbf{x}^H[n]]$  and

$$\mathbf{d} = \text{cum}\{e[n] : p, e^*[n] : q - 1, \mathbf{x}^*[n]\}, \quad p + q \geq 3 \quad (11)$$

The SEA is a computationally efficient algorithm with fast convergence speed (in terms of ISI) but no guarantee of convergence for finite SNR and data.

### 2.4. IFC

The IFC( $p, q$ ) [3] find the optimum  $\mathbf{v}$  by maximizing the following criteria:

$$J_{p,q}(\mathbf{v}) = \frac{|C_{p,q}^e|}{[C_{1,1}^e]^{(p+q)/2}}, \quad p + q \geq 3 \quad (12)$$

which is a highly nonlinear function of  $v[n]$  without a closed-form solution for the optimum  $\mathbf{v}$ . Chi, Feng and Chen [4] proposed a fast gradient type iterative algorithm as follows:

#### Algorithm 1:

At the  $i$ th iteration,  $\mathbf{v}^{[i]}$  is obtained through the following two steps.

(T1) Update  $\hat{\mathbf{v}}$  using (10) with  $e[n] = e^{[i-1]}[n]$  used in  $\mathbf{d}$  (see (11)), and obtain the associated  $e^{[i]}[n]$ .

(T2) If  $J_{p,q}(\hat{\mathbf{v}}) > J_{p,q}(\mathbf{v}^{[i-1]})$ , update  $\mathbf{v}^{[i]} = \hat{\mathbf{v}}$ , otherwise update  $\mathbf{v}^{[i]}$  by

$$\mathbf{v}^{[i]} = \mathbf{v}^{[i-1]} + \mu \text{sgn}(C_{p,q}^u) \hat{\mathbf{v}} \quad (13)$$

such that  $J_{p,q}(\mathbf{v}^{[i]}) > J_{p,q}(\mathbf{v}^{[i-1]})$ , and obtain the associated  $e^{[i]}[n]$ .

Algorithm 1 requiring real  $x[n]$ , or complex  $x[n]$  and  $p = q$ , shares the computational efficiency and convergence speed of the SEA with guaranteed convergence.

### 3. EQUIVALENCE OF SEA(2,2), IFC(2,2) AND CMA

Chi, Feng and Chen [4] have proven the following fact:

(F3) SEA( $p, q$ ) and IFC( $p, q$ ) are equivalent as  $x[n]$  is real and  $p + q \geq 3$  or as  $x[n]$  is complex and  $p = q \geq 2$ .

As mentioned in (F2),  $\mathbf{x}[n]$  and  $\mathbf{b}[n]$  are causally invertible. Therefore, deconvolution with  $\mathbf{x}[n]$  is equivalent to deconvolution with  $\mathbf{b}[n]$ . Let

$$e[n] = \mathbf{v}^T \mathbf{b}[n] \quad (14)$$

Replacing  $\mathbf{x}[n]$  and  $\mathbf{R}_x$  in (10) with  $\mathbf{b}[n]$  and  $\mathbf{R}_b$ , respectively, and replacing  $e[n]$  in (11) with the one given by (14) for  $p = q = 2$  through some simplification yields

$$\hat{\mathbf{v}} = \mathbf{R}_b^{-1} E[|e[n]|^2 e[n] \mathbf{b}^*[n]] \quad (15)$$

except for a scale factor. On the other hand, substituting (14) into  $J_{\text{CM}}(\mathbf{v})$  given by (9), one can easily show that the optimum  $\hat{\mathbf{v}}$  associated with the  $J_{\text{CM}}(\mathbf{v})$  is the same as the one given by (15) except for a scale factor. Therefore, we have shown the following theorem:

**Theorem 1.** Both SEA( $p, q$ ) with  $p = q = 2$  and CMA are equivalent.

By (F3) and Theorem 1, we have the following fact:

(F4) The CMA, IFC( $p, q$ ) and SEA( $p, q$ ) are equivalent as  $p = q = 2$ . Therefore, they share some properties reported in [4–7], such as perfect equalization property and relation to nonblind minimum mean square error (MMSE) equalizer.

#### 4. LATTICE IMPLEMENTATIONS

Let us present lattice implementations for SEA( $p, q$ ), IFC( $p, q$ ) and CMA only for the case of  $p = q = 2$  below.

##### 4.1. Off-Line Processing

Feng and Chi have reported two off-line lattice SEA (LSEA) [9] using  $\mathbf{b}_k[n]$  and  $\mathbf{f}_k[n]$ , respectively. Next, let us present two lattice implementations for IFC that are modifications of Algorithm 1 with  $\mathbf{x}[n]$  replaced by  $\mathbf{b}_k[n]$  and  $\mathbf{f}_k[n]$ , respectively.

**LIFC-B Algorithm:** At the  $i$ th iteration,  $\mathbf{v}^{[i]}$  is obtained through the following two steps.

- (S1) Compute  $\hat{\mathbf{v}}$  by (15) where  $e[n] = e^{[i-1]}[n]$  is obtained by (14) at the  $(i-1)$ th iteration.
- (S2) If  $J_{2,2}(\hat{\mathbf{v}}) > J_{2,2}(\mathbf{v}^{[i-1]})$ , update  $\mathbf{v}^{[i]} = \hat{\mathbf{v}}$ , otherwise update  $\mathbf{v}^{[i]}$  through a gradient-type optimization procedure with the gradient

$$\nabla J_{2,2} \propto \text{sgn}(C_{2,2}^u) \mathbf{R}_b(\hat{\mathbf{v}} - \mathbf{v}^{[i-1]}). \quad (16)$$

**LIFC-F Algorithm:** Let

$$e[n] = \mathbf{v}^T \mathbf{f}_k[n] \quad (17)$$

where  $k$  is sufficiently large such that (F1) applies to  $\mathbf{f}_k[n]$ . At the  $i$ th iteration,  $\mathbf{v}^{[i]}$  is obtained through the same procedure as the previous LIFC-B algorithm except that  $\mathbf{b}[n]$  and  $\mathbf{R}_b$  are replaced by  $\mathbf{f}_k[n]$  and  $\mathbf{R}_{f_k}$ , respectively, with  $e^{[i]}[n]$  obtained by (17) and  $\nabla J_{2,2}$  obtained by

$$\nabla J_{2,2} \propto \text{sgn}(C_{2,2}^u)(\hat{\mathbf{v}} - \mathbf{v}^{[i-1]}). \quad (18)$$

A worthy remark regarding the proposed LIFC-B and LIFC-F algorithms is as follows:

(R1) The proposed LIFC-B and LIFC-F algorithms are computationally efficient (without need of matrix inversion) with guaranteed convergence, whereas the latter converges faster than the former since  $\mathbf{f}_k[n]$  approximates an amplitude equalized signal by (F1).

(R2) As deriving the LIFC-B and LIFC-F algorithms (maximizing  $J_{2,2}$ ), one can readily obtain two lattice CMA algorithms (minimizing  $J_{\text{CM}}$ ), using  $\mathbf{b}_k[n]$  and  $\mathbf{f}_k[n]$ , respectively, that also share the implementation merits of the LIFC-B and LIFC-F algorithms mentioned in (R1).

##### 4.2. Adaptive Processing

Let  $\mathbf{v}_n$  denote the estimate of  $\mathbf{v}$  as  $\mathbf{x}[n]$  is processed. An adaptive SEA reported in [1] is as follows:

$$\mathbf{v}_{n+1} = \mathbf{v}_n + \mu \mathbf{Q}_{n+1} \mathbf{x}^*[n+1] e[n] (\gamma - |e[n]|^2) \quad (19)$$

$$\mathbf{v}_{n+1} = \mathbf{v}_{n+1} / \|\mathbf{v}_{n+1}\| \quad (20)$$

where  $\mu$  is the step size parameter, and

$$e[n+1] = \mathbf{v}_{n+1}^T \mathbf{x}[n+1] \quad (21)$$

$$\mathbf{Q}_{n+1}^{-1} = (1 - \mu) \mathbf{Q}_n^{-1} + \mu \mathbf{x}[n+1] \mathbf{x}^H[n+1] \quad (22)$$

With  $\mathbf{Q}_{n+1}$  and  $\mathbf{x}[n]$  in (19) replaced by  $\mathbf{R}_{f_k}^{-1}$  and  $\mathbf{f}_k[n]$ , one can obtain

$$\mathbf{v}_{n+1} = \mathbf{v}_n + \mu \{\gamma - |e[n]|^2\} e[n] \mathbf{f}_k^*[n+1] \quad (23)$$

$$e[n+1] = \mathbf{v}_{n+1}^T \mathbf{f}_k[n+1] \quad (24)$$

**Lattice SE-IF-CM Algorithm:** For each  $x[n+1]$ , two signal processing steps are performed as follows:

- (U1) Obtain  $\mathbf{f}_k[n+1]$  by processing  $x[n+1]$  with the adaptive least squares lattice (LSL) LPE filter [8].
- (U2) Update  $\mathbf{v}_{n+1}$  and  $e[n+1]$  using (23) and (24), respectively.

Two worthy remarks regarding the lattice SE-IF-CM algorithm are as follows:

(R3) The lattice SE-IF-CM algorithm is exactly a lattice CMA algorithm since (U2) is the same as the adaptive CMA [2] and an adaptive IFC algorithm [3] except that  $\mathbf{f}_k[n]$  is replaced with  $\mathbf{x}[n]$ .

(R4) The proposed lattice SE-IF-CM algorithm with low computational load (without matrix multiplication operations) converges faster than the adaptive SEA given by (19) through (22) and the adaptive CMA since the adaptive LSL algorithm in (U1) performs as a fast amplitude equalizer.

## 5. SIMULATION RESULTS

Two examples are presented to support our analytic results and the lattice structure based algorithms.

### Example 1: Off-line Processing

The source signal  $u[n]$  was assumed to be a 4-QAM signal with unity variance and a real channel  $h[n]$  was taken from [1] as plotted in Figure 1(a). The equalizer  $v[n]$  was assumed to be a causal FIR filter of order  $L = 50$ . Thirty independent runs for data length  $N = 4096$  and SNR = 20 dB (complex white Gaussian noise) were performed using CMA and SEA(2,2) with the initial condition  $v[n] = \delta[n-L/2]$ , respectively. The averages of thirty independent estimates of equalizer  $v[n]$  obtained using CMA and SEA(2,2) are displayed in Figures 1(b) and 1(c), respectively, where only equalizer real parts are shown since imaginary parts are almost zero. These results justify Theorem 1.

Moreover, Algorithm 1, LIFC-B and LIFC-F algorithms and a gradient-based IFC algorithm were also employed to process the same simulation data. Figure 2 shows the average of the thirty  $J_{2,2}$ 's with respect to iteration number associated with LIFC-F ( $k=50$ ) algorithm (dash line), LIFC-B algorithm (dash-dotted line), Algorithm 1 (dotted line) and the gradient-based IFC algorithm (solid line). Figure 2 depicts that the LIFC-F algorithm and Algorithm 1 converge faster than the other two algorithms (see (R1)) and the gradient-based IFC algorithm converges slower than all the other algorithms. These simulation results support the efficacy of the proposed LIFC-B and LIFC-F algorithms.

### Example 2: Adaptive Processing

The source signal  $u[n]$  was assumed to be a 2-PAM (+1, -1) signal. The same channel  $h[n]$  as shown in Figure 1(a) was used, and SNR = 20 dB (real white Gaussian noise). Figure 3 shows some simulation results (average of thirty independent ISI's versus iteration number) for  $L = 24$  using the adaptive SEA with  $p = q = 2$  and  $\mu = 0.0026$ , the adaptive CMA with  $\mu = 0.00215$  and the proposed adaptive lattice SE-IF-CM algorithm with  $k = 24$  and  $\mu = 0.002$ . Note that the value of the step size  $\mu$  used by each adaptive algorithm was chosen through some trial-and-errors such that its performance is "best" in terms of convergence speed and ISI. One can see, from Figure 3, that the proposed adaptive lattice SE-IF-CM algorithm (solid line) converges faster than the other two adaptive algorithms with ISI slightly smaller than those associated with the other two adaptive algorithms. These simulation results justify the efficacy of the proposed adaptive lattice SE-IF-CM algorithm (see (R4)).

## 6. CONCLUSIONS

We have shown the equivalence of the CMA, SEA( $p, q$ ) and IFC( $p, q$ ) for  $p = q = 2$  as presented in Theorem 1 and (F4), and therefore, any performance analyses for one of them apply to the others. Furthermore, two computationally efficient off-line processing algorithms, LIFC-F and LIFC-B algorithms for  $p = q = 2$  were presented, while the former is preferable to both the latter and Chi, Feng and Chen's Algorithm 1 due to faster convergence (see (R1)). For adaptive processing, a computationally efficient lattice SE-IF-CM algorithm for  $p = q = 2$  was presented that has computational complexity similar to the adaptive CMA and converges faster than both the adaptive SEA and the adaptive CMA with similar resultant ISI (see (R3) and (R4)). The efficacy of the proposed adaptive lattice SE-IF-CM algorithm and the proposed analytic results were supported by some simulation results. As a final remark, for  $p \neq q$  or  $p = q \neq 2$ , lattice implementations of the SEA and IFC can be similarly developed.

## REFERENCES

- [1] O. Shalvi and E. Weinstein, "Super-exponential methods for blind deconvolution," *IEEE Trans. Information Theory*, vol. 39, no. 2, pp. 504-519, March 1993.
- [2] J.R. Treichler and B.G. Agee, "A new approach to multipath correction of constant modulus signals," *IEEE Trans. Acoustics, Speech and Signal Processing*, vol. 31, no. 2, pp. 349-472, Apr. 1983.
- [3] O. Shalvi and E. Weinstein, "New criteria for blind deconvolution of nonminimum phase systems (channels)," *IEEE Trans. Information Theory*, vol. 36, pp. 312-321, March 1990.
- [4] C.-Y. Chi, C.-C. Feng and C.-Y. Chen, "Performance of super-exponential algorithm for blind equalization," *Proc. IEEE VTC2000-Spring*, Tokyo, Japan, May 15-18, 2000, pp. 1864-1868.
- [5] H.H. Zeng, L. Tong and C.R. Johnson, "Relationships between the constant modulus and Wiener receivers," *IEEE Trans. Signal Processing*, vol. 44, no. 4, pp. 1523-1538, July 1998.
- [6] C.-C. Feng and C.-Y. Chi, "Performance of cumulant based inverse filters for blind deconvolution," *IEEE Trans. Signal Processing*, vol. 47, no. 7, pp. 1922-1935, July 1999.
- [7] C.-C. Feng and C.-Y. Chi, "Performance of Shalvi and Weinstein's deconvolution criteria for channels with/without zeros on the unit circle," *IEEE Trans. Signal Processing*, vol. 48, no. 2, pp. 571-575, Feb. 2000.
- [8] S. Haykin, *Adaptive Filter Theory*, 2nd Ed., Prentice-Hall, Upper Saddle River, New Jersey, 1991.
- [9] C.-C. Feng and C.-Y. Chi, "A two-step lattice super-exponential algorithm for blind equalization," *Proc.*

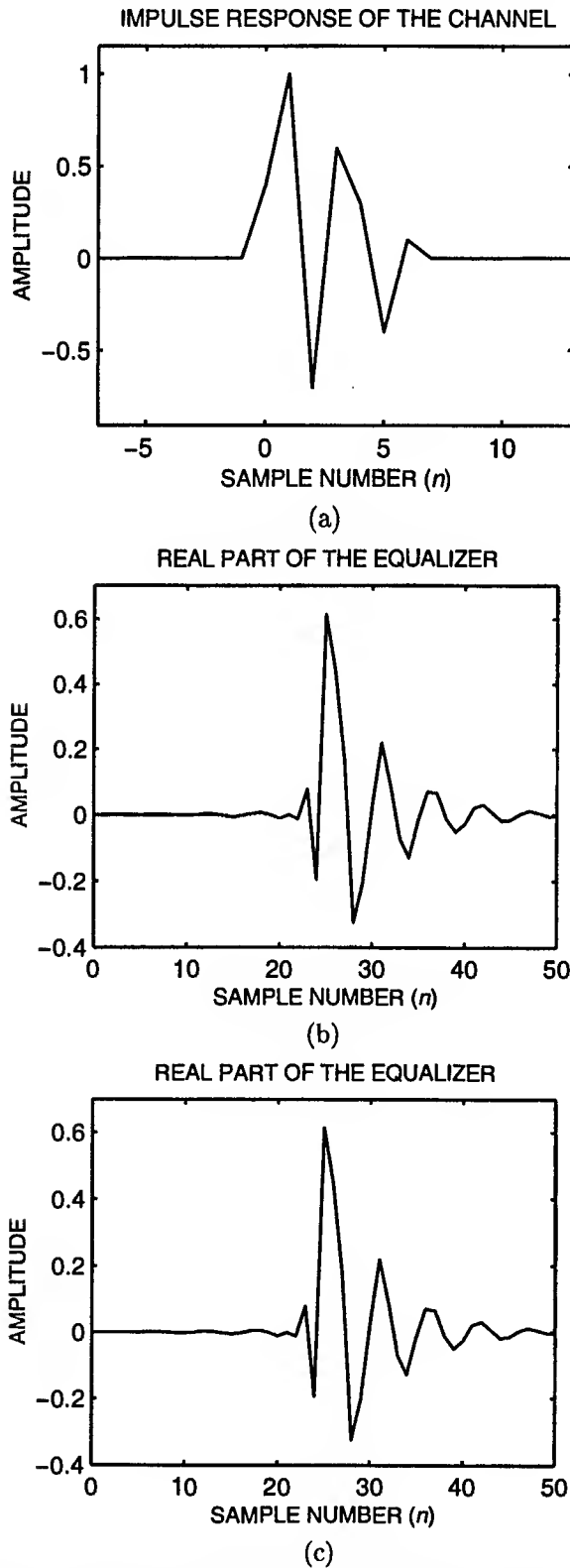


Figure 1. Simulation results for  $N = 4096$  and  $\text{SNR} = 20$  dB. (a) The channel impulse response; (b) average of thirty

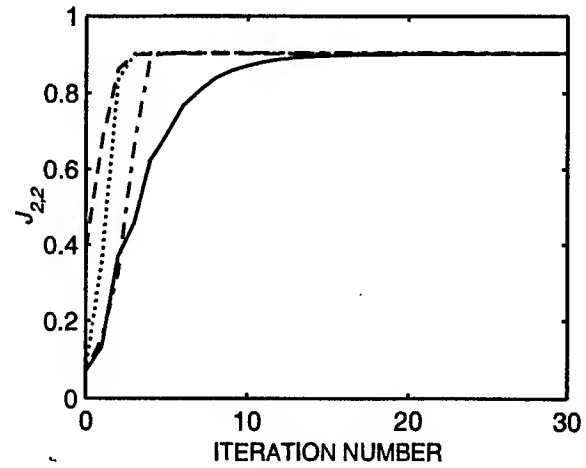


Figure 2. Average of thirty  $J_{2,2}$ 's associated with LIFC-F ( $k=50$ ) (dash line) algorithm, LIFC-B algorithm (dash-dotted line), Algorithm 1 (dotted line) and the gradient-based IFC algorithm (solid line).

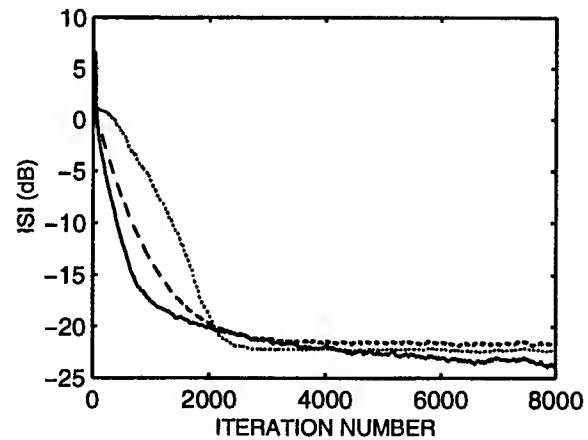


Figure 3. Simulation results (ISI versus iteration number) using the adaptive CMA (dash line) with  $\mu = 0.00215$ , the adaptive SEA (dotted line) with  $p = q = 2$  and  $\mu = 0.0026$  and the proposed adaptive lattice SE-IF-CM algorithm (solid line) with  $k = 24$  and  $\mu = 0.002$ , respectively.



# AN EFFICIENT ALGORITHM FOR GAUSSIAN-BASED SIGNAL DECOMPOSITION

Zou Hong, student member, *IEEE* and Bao Zheng, member, *IEEE*  
(P.O.Box 374, Xidian University, Xi'an, P.R.China 710071)

## ABSTRACT

Parameter estimation of each atom in the adaptive Gaussian basis representation (AGR) is a key problem determining signal decomposition results, which requires an effective parameter estimation algorithm. In this paper, an efficient algorithm is proposed to estimate the time-frequency atoms in AGR. Compared with the algorithm presented in AGR, the parameter estimation is more adaptive, and the estimation accuracy is greatly improved through an iterative method. Numerical simulations confirm the results.

## 1. INTRODUCTION

The adaptive Gaussian basis representation (AGR)[1], presented by Qian and Chen, is to decompose the analyzed signal into a linear expansion of Gaussian atoms, in which the parameters of the Gaussian atoms can be adjusted to best match the signal, and then a crossterms free time-frequency distribution is obtained. A similar decomposition result is also obtained by S.Mallat[2] with waveforms. Although the decomposition error is monotonically decreasing, the results still rely heavily on the parameter estimation algorithm. In this paper, a new efficient algorithm to compute the Gaussian atoms in AGR is proposed, which avoids determining the searching intervals of Gaussians' time centers and width in advance, and greatly improves the estimation accuracy.

## 2. ALGORITHM DESCRIPTIONS

### 2.1 Brief introduction of AGR

The goal of AGR is to represent a signal of

interest  $s(t)$  by a class of localized time-frequency

atoms  $h_k(t)$ :

$$s(t) = \sum_{k=0}^M C_k h_k(t) \quad (1)$$

Where,

$$h_k(t) = \left(\frac{\alpha_k^2}{\pi}\right)^{1/4} \cdot \exp\left\{-\frac{\alpha_k^2(t-t_k)^2}{2}\right\} e^{j\omega_k(t-t_k)} \quad (2)$$

$h_k(t)$  is a normalized Gaussian,  $\alpha_k$  determines the width of the Gaussian function, and  $(t_k, \omega_k)$  is its time-frequency center. Estimation of  $h_k(t)$  at the  $k$ th stage is equivalent to:

$$|C_k|^2 = \max_{\alpha_k, t_k, \omega_k} |\langle s_k(t), h_k(t) \rangle|^2 \quad (3)$$

and  $s_k(t)$  is the remainder of the orthogonal projection of  $s_{k-1}(t)$  onto  $h_{k-1}(t)$ .

### 2.2 Adaptive algorithm for Gaussian parameters estimation

Assume it is at the  $k$ th stage in AGR, and the remainder is  $s_k(t)$ . The procedure of the new proposed algorithm to estimate  $\alpha_k, t_k, \omega_k$  and  $C_k$  is presented as follows. Let  $n = 0$ .

**Step one:** Estimate  $\omega_k$  and initial value  $t_{kn}(n=0)$  of

$t_k$  with  $s_k(t)$  by use of Spectrogram:

$$(\omega_k, t_{kn}) = \operatorname{argmax}_{\omega, t} \left\{ \left| \int s_k(\tau) g^*(\tau - t) e^{-j\omega\tau} d\tau \right|^2 \right\} \quad (4)$$

Spectrogram is used here for the following reasons:

1) Although Spectrogram is low in time-frequency resolution, it can avoid crossterms, which is helpful in extracting initial values. 2) Spectrogram of the input signals displays parallel lines on the time-frequency plane, thus peak searching on this plane can decide the time-frequency center of each component, while it is not the case for frequency modulated signals. 3) if the weighting window  $g(t)$  is long enough, estimation of

$\omega_k$  will be satisfactorily accurate for subsequent use.

Hence in the following procedure, we assume that  $\omega_k$  is equal to the real value.

**Step two:** Estimate initial value  $\alpha_{kn}(n=0)$  of  $\alpha_k$ .

Since  $s_k(t)$  and  $h_k(t)$  are Gaussian-shaped, and FFT of both are symmetrical about  $\omega = \omega_k$ , the method adaptively estimating the width of the time-frequency kernel[3] is used here for choosing  $\alpha_{kn}$ .

That is:

$$\alpha_{kn} = \max \{ \alpha_i, (i = 1, 2, \dots, l) \mid R_{k,i}(\alpha) < \eta \} \quad (5)$$

Where

$$R_{k,i}(\alpha) = \left| \int \frac{d}{d\omega} [ |S_k(\omega) \cdot G_{k,i}^*(\omega)| ] d\omega \right| \quad (6)$$

$$g_{k,i}(t) = \exp \left\{ -\frac{\alpha_{k,i}^2 (t - t_{kn})^2}{2} \right\} e^{j\omega_k(t - t_{kn})} \quad (7)$$

$G_{k,i}(\omega)$  and  $S_k(\omega)$  are normalized Fourier

Transform (FFT) of  $g_{k,i}(t)$  and  $s_k(t)$ , respectively.

$\alpha_{k,i}$  ( $i = 1, 2, \dots, l$ ) is  $l$  different values roughly chosen

for  $\alpha_{kn}$  design, and  $\eta$  is a threshold to control the range of  $\alpha_{kn}$ .

**Step three:** Accurately estimate  $\alpha_k$  from  $\alpha_{kn}(n=0)$ .

$$\text{Let } s_k(t) = \exp \left\{ -\frac{\alpha_0^2 (t - t_0)^2}{2} \right\} e^{j\omega_0(t - t_0)}$$

$$\text{and } h_k(t) = \exp \left\{ -\frac{\alpha_{kn}^2 (t - t_{kn})^2}{2} \right\} e^{j\omega_k(t - t_{kn})}$$

respectively, and  $S_k(\omega)$  and  $H_k(\omega)$  are

normalized FFT of  $s_k(t)$  and  $h_k(t)$ . Then we have:

$$P_0 = \int |S_k(\omega)|^2 d\omega = \sqrt{\frac{\pi}{2}} \alpha_0 \quad (8)$$

$$P_{kn} = \int |S_k(\omega) H_k^*(\omega)| d\omega = \frac{\sqrt{\pi} \alpha_0 \alpha_{kn}}{\sqrt{\alpha_0^2 + \alpha_{kn}^2}} \quad (9)$$

Thus the following result can be obtained:

$$\begin{cases} P_{kn} > P_0 & \text{when } \alpha_{kn} > \alpha_0 \\ P_{kn} < P_0 & \text{when } \alpha_{kn} < \alpha_0 \end{cases} \quad (10)$$

Therefore, starting from the initial value  $\alpha_{kn}, P_{kn}$ , and

take  $P_0$  as a reference value,  $\alpha_{kn}$  can be adjusted

adaptively according to the difference between  $P_{kn}$  and

$P_0$  until  $\alpha_{kn} = \alpha_0$ :

$$\alpha_{kn+1} = \alpha_{kn} + u(P_0 - P_{kn}) \quad (n = 0, 1, 2, 3, \dots)$$

Where  $u > 0$  is the convergence factor. In practical applications, however, it is very difficult to choose  $u$ , since Eq.(8),(9) do not hold exactly, which is resulted from the limited length and discrete sampling of the analyzed signal. Consequently, we adopt the following numerical algorithm:

1. Because the initial value  $\alpha_{kn}(n=0)$  is always

much larger than the real value, namely,  $P_{k0} > P_0$

is always true, we start from  $n=1$ , let

$\alpha_{kn} = 2^{-1} \alpha_{kn-1}$  until  $P_{kn} < P_0$ . Assume  $n=q$  at this time.

2. Adjust  $\alpha_{kn}$  with the rule below until

$$|P_{kn} - P_0| < \varepsilon, \text{ and } n=q+1, q+2, \dots:$$

$$\alpha_{kn} = \text{mid}\{\alpha_{kn-1}, \min[\alpha_{ki}, i=1, n-2 | \alpha_{ki} > \alpha_{kn-1}]\}$$

$$\text{if } P_{kn} < P_0$$

$$\alpha_{kn} = \text{mid}\{\alpha_{kn-1}, \max[\alpha_{ki}, i=1, n-2 | \alpha_{ki} < \alpha_{kn-1}]\}$$

$$\text{if } P_{kn} > P_0$$

Where  $\varepsilon$  is a given threshold controlling estimation precision of  $\alpha_k$ , and  $\text{mid}\{x, y\}$  means the mid value between  $x$  and  $y$ . Thus we can get a more accurate estimation of  $\alpha_k$ .

**Step four:** Accurately estimate  $t_k, C_k$  with the estimated  $\alpha_k$  and  $\omega_k$  according to Eq.(3):

$$|C_k|^2 = \max_{t_k} |< s_k(t), h_k(t, \alpha_k, \omega_k) >|^2 \quad (11)$$

### 3. IMPLEMENTATION CONSIDERATIONS

The following are observations needed to be

considered in practical implementations:

1. When multicomponents exist in the analyzed signals, long integral range used in Eq. (8) and (9) may destroy the relationship in Eq.(11). Hence, the 3dB bandwidth around the spectrum peak of each component is adopted as the integral range of Eq. (8), (9).
2. When multicomponents exist in the signal, great errors will occur when estimating  $C_k$  and  $t_k$  with Eq.(11). An iterative algorithm similar to the "RELAX" method[4] can be used to improve the accuracy. That is, each time after a new component is estimated, all previously computed components are re-estimated again.

### 4. SIMULATIONS

To demonstrate the effectiveness of the algorithm presented above, we apply it to a synthetic signal composed of four Gaussian components. The signal is 256-point,  $\varepsilon = 0.001$ , and the time center of each component is represented by the sampling index. Comparison of the estimated parameters and the true values is given in Table 1, which shows high accuracy of the developed algorithm. The signal energy of the residual is less than 0.03 of that contained in the original signal.

Fig.1(a),(b) show the contour plot of Spectrogram and Wigner-Ville distribution(WVD) of the synthetic signal above, respectively. Fig.1(c) is the WVD of the signal after decomposition. It can be seen comparing (b) and (c) that, by use of the signal decomposition method, crossterms in WVD can be eliminated effectively, and the signal's time-frequency resolution is high.

### 5. CONCLUSION

Atoms estimation in signal decomposition method determines the decomposition effects and the description of signals. The authors present an efficient numerical algorithm to improve the parameter estimation accuracy in AGR. The advantage is : The procedure is simple; The computation is not complex; Estimation of all

parameters of the Gaussian atoms is more adaptive; The estimation precision of  $\alpha_k$  and  $t_k$  is improved greatly via an iterative adaptive method.

## ACKNOWLEDGEMENT

This work is supported by the NSF of China.

## REFERENCES

- [1] Shie Qian and Dapang Chen, 'Signal representation using adaptive normalized Gaussian function', Signal Processing 36 (1994), pp.1-11.
- [2] Stephane G. Mallat and Zhifeng Zhang, 'Matching pursuits with time-frequency dictionaries', IEEE Trans. SP. Vol.41, NO. 12, pp.3397-3415,1991.
- [3] Zou Hong and Bao Zheng, 'An adaptive-kernel design method based on ambiguity domain', Proceedings of IEEE-SP International Symposium on Time-Frequency and Time-Scale Analysis, pp.197-200,1998.
- [4] Jian Li and Petre Stoica, 'Efficient mixed-spectrum estimation with applications to target feature extraction', IEEE Trans. SP. Vol. 44, NO.2, pp.281-295,1996

**Table 1:** Comparison between the estimated and true values

		$C_k$	$t_k$	$f_k$	$\alpha_k$
1	T	2	100	2.5	0.056
	E	2.0073	100	2.5	0.0562
2	T	2.0	66	0.77	0.07
	E	2.0027	66	0.7682	0.0699
3	T	1.0	80	1.6	0.033
	E	1.0093	80	1.6016	0.033
4	T	1.0	162	5.0	0.1
	E	1.0046	162	5.0	0.1003

T: Denotes the true value.

E: Denotes the estimated values.

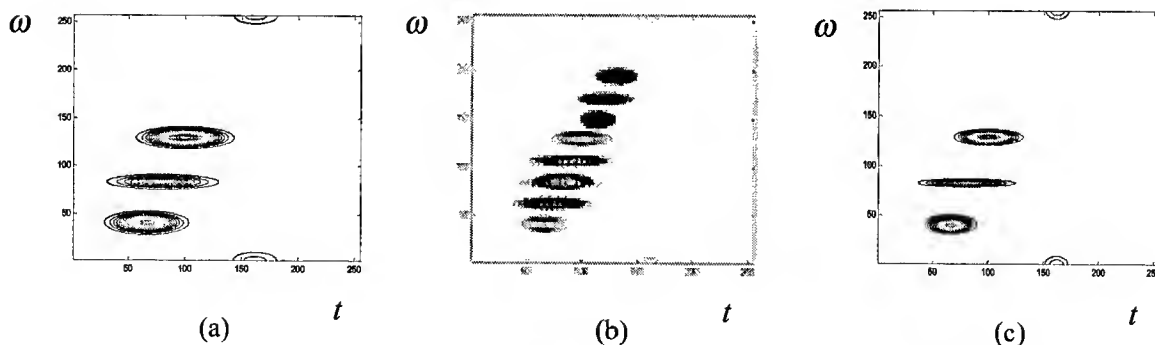


Fig.1 Contour plot of (a) Spectrogram (b) WVD (c) WVD of decomposed signal

# CONSISTENT ESTIMATION OF SIGNAL PARAMETERS IN NON-STATIONARY NOISE

J. Friedmann, E. Fishler and H. Messer

Department of Electrical Engineering-Systems, Tel Aviv University,  
Tel Aviv 69978, Israel,

E-mail: : {jonaf, eranf, messer}@eng.tau.ac.il.

## ABSTRACT

This paper addresses the problem of estimating the parameters of a deterministic signal in non stationary, white, Gaussian noise. It is proposed to model the time varying white Gaussian noise with an unknown deterministic variance sequence that changes every sample. While making relatively little assumptions on the non stationarity of the noise, this type of modeling gives rise to different difficulties. In the paper we identify the resulting difficulties and discuss possible solutions.

## 1. INTRODUCTION

The problem of estimating the parameters of a deterministic signal in additive noise is a problem of many disciplines. Numerous amount of applications exist for this model and range from underwater acoustics to cellular communications. In this paper we take interest in the case where the additive noise is white, non stationary and Gaussian. First, a novel approach for modeling the non stationarity is examined. Then, for this modeling, specific estimation algorithms are proposed and analyzed.

The observed time series at instant  $n$ ,  $y_n$ , is modeled as follows:

$$y_n = s_n(\theta) + v_n. \quad (1)$$

where  $s_n(\theta)$  is a deterministic signal which is known to within a parameter vector  $\theta$ , and  $v_n$  is a white, non stationary, Gaussian noise sequence.

The model in (1) represents many important applications for which there exists abundant literature. For each application, various signals are used and different assumptions are made on the noise distribution. The present paper relaxes the traditional assumption on the stationarity of the noise.

Most of the works in the literature dealing with non stationary processes are based on a specific, or a parametric model for the non stationarity. The proposed model is then justified for the problem of interest. In [6], for instance, a parametric approach is proposed to characterize a non stationary process using a time dependent *ARMA* process.

However, since miss-modeling may cause significant errors, it would sometimes be best to avoid making any assumptions on the characteristics of the non stationarity, and to model the non stationary process in a more general way. A natural way is to model the noise variance sequence in (1) as a sequence of positive, deterministic, unknown parameters. However, this type of modeling causes various difficulties.

The main difficulty in such modeling arises since the number of unknown parameters increases with the number of samples. The family of problems which are characterized by this property was first presented by Neyman and Scott [7]. Their goal was to find "Consistent estimates based on partially consistent observations". In their classic paper they show that sometimes maximum likelihood (*ML*) estimation fails to provide consistent estimates. They also give examples in which *ML* estimates are consistent but not optimal in the mean square error sense. We follow their definitions to present the problem of interest.

In model (1) the parameters of interest are the signal parameters while the unknown, time varying noise variances are the nuisance parameters. The signal parameters,  $\theta$ , are called *structural* parameters. As the number of observed data samples approaches infinity, these parameters affect the probability law of an infinite amount of data samples. The noise parameters,  $\sigma_1^2, \dots, \sigma_N^2$  are called *incidental* parameters. These parameters affect the probability law of a finite amount of data samples even when the number of observed data samples approaches infinity. For the model presented in (1) it is shown in [3] that *ML* estimates for  $\theta$  do not exist in general and therefore, other solutions should be looked for.

## 2. OPTIMALITY CRITERIA

In estimation problems in which there exist only structural parameters, under mild regularity conditions, *ML* estimator is the asymptotic, universal, optimal estimator in the mean square error sense, *i.e.*, regardless of the values of the parameters to be estimated, *ML* estimation achieves asymptotically the best performance possible described by the Cramer Rao bound (*CRB*). However, in cases where

the estimation problem includes incidental parameters, the *CRB* is not always attainable and therefore may serve only as a lower bound. Thus, an estimator may be optimal without attaining the *CRB*.

In the search for the optimal estimator (in the mean square sense), few options rise. One would very much like to find a universal optimal estimator in the presented family of estimation problems. In other words, to find an estimator for the structural parameters, which is optimal for any given sequence of incidental parameters. However, this is in general impossible. It is possible, however, to find an optimal estimator in a restricted family of estimators. Such a criterion of optimality is discussed in [1]. A different optimality criterion is to obtain the optimal estimator when the unknown incidental parameters is regarded as an i.i.d. sample from an unknown, but fixed distribution belonging to a non parametric family of distributions. Such an estimating problem is said to be semi-parametric and an optimal estimator may be derived for different problems using adaptive estimation technique (see [2]). Finally, another possible optimality criterion is to obtain the optimal estimator in a minimax sense. That is, to minimize the estimating loss provided the incidental parameters attain the worst sequence of values (see [4]).

### 3. THE CRAMÉR RAO BOUND

For the problem of estimating the parameters of a deterministic signal in non stationary noise (1), the Fisher Information matrix (*FIM*) is a block diagonal matrix:

$$I(\theta, \sigma_1 \dots \sigma_N^2) = \begin{bmatrix} I(\theta) & 0 \\ 0 & I(\sigma_1 \dots \sigma_N^2) \end{bmatrix}. \quad (2)$$

Therefore, the *CRB* for the structural parameters is

$$CRB(\theta) = I^{-1}(\theta) = \left( \sum_{i=1}^N \frac{1}{\sigma_n^2} \frac{\partial s_n(\theta)}{\partial \theta} \frac{\partial s_n(\theta)^T}{\partial \theta} \right)^{-1}. \quad (3)$$

As shown in [3], for the problem of interest, *ML* estimates do not exist. Therefore, the *CRB* does not describe the asymptotic covariance of the *ML* estimates. However, it may serve as a lower bound for any estimator.

### 4. GENERALIZED M-ESTIMATES

In the following we study estimating equations depending on the parameters of interest only. We consider estimates of the signal parameters,  $\theta$ , obtained by minimizing a cost function consisting of a sum of the time varying functions,  $\psi_n(\cdot)$ :

$$\hat{\theta} = \arg \min_{\theta} \sum_{i=1}^N \psi_n(y_n - s_n(\theta)). \quad (4)$$

This is a time varying extension of the *M* estimates, first presented by Huber [5]. If the sequence  $\sigma_1^2, \dots, \sigma_N^2$  were known the cost function may be chosen to be the likelihood function. Other functions may also be considered. Differentiating  $\psi_n(\cdot)$  with respect to  $\theta$  the estimates are obtained as solutions of the estimating equations,

$$\sum_{i=1}^N \varphi_n(y_n - s_n(\hat{\theta})) \frac{\partial s_n(\hat{\theta})}{\partial \theta} = 0, \quad (5)$$

where  $\varphi_n(\cdot) = \psi'_n(\cdot)$ . If the cost function is chosen to be the log likelihood function, then the estimating function is the associated score function. As suggested by Huber [5], the choice of the nonlinear function  $\varphi_n$  should be guided not only by performance considerations but also by considerations of robustness. Huber suggested robustness to deviations from the assumed noise distribution. Instead, we suggest the choice of estimating functions robust to the unknown non-stationarity, *i.e.*, the unknown sequence  $\sigma_1^2, \dots, \sigma_N^2$ .

We now turn to compute the asymptotic performance of the estimates as a function of the estimating functions  $\varphi_n$ .

#### 4.1. Small error analysis

The small error behavior of the estimates is determined from a first order expansion of the estimating equations (5) and is derived in [3]. Under mild conditions, it can be shown that for small errors:

$$(\hat{\theta} - \theta) \approx A^{-1} \cdot \sum_{i=1}^N \varphi_n(v_n) \frac{\partial s_n(\theta)}{\partial \theta} \quad (6)$$

where  $A$  is defined by

$$A = \sum_{i=1}^N E \varphi'_n(v_n) \frac{\partial s_n(\theta)}{\partial \theta} \frac{\partial s_n(\theta)^T}{\partial \theta^T}. \quad (7)$$

Each term in the stochastic expansion (6) depends on a single value of  $v_n$  so that it is a sum of independent terms because the noise sequence itself is modeled as independent. It is then clear, that if  $E \varphi_n(v_n) = 0$  then the small error bias is zero. Moreover, the covariance of the small estimation error is then given by:

$$\text{Cov}(\hat{\theta}) = A^{-1} \left( \sum_{i=1}^N E \varphi_n^2(v_n) \frac{\partial s_n(\theta)}{\partial \theta} \frac{\partial s_n(\theta)^T}{\partial \theta} \right) A^{-1}. \quad (8)$$

#### 4.2. Example

One example for a generalized *M* type estimator (for the structural parameters) is the weighted least squares estima-

tor (*WLS*). In this case the cost function is given by:

$$\sum_{i=1}^N \left( \frac{y_n - s_n(\hat{\theta})}{k_n} \right)^2, \quad (9)$$

i.e.,  $\psi_n(\cdot) = (\frac{\cdot}{k_n})^2$  where  $k_n$  are predetermined weights. The estimating functions are given by the following set of nonlinear equations:

$$\sum_{i=1}^N \left( \frac{y_n - s_n(\hat{\theta})}{k_n^2} \right) \frac{\partial s_n(\hat{\theta})}{\partial \theta} = 0, \quad (10)$$

i.e.,  $\varphi_n(x) = \frac{x}{k_n^2}$ , so  $\varphi'_n = \frac{1}{k_n^2}$ , and the asymptotic covariance of the this  $M$  type estimator is given by:

$$\text{Cov}(\hat{\theta}) = \left( \sum_{i=1}^N \frac{1}{k_n^2} \frac{\partial s_n(\theta)}{\partial \theta} \frac{\partial s_n(\theta)^T}{\partial \theta} \right)^{-1} \quad (11)$$

$$\left( \sum_{i=1}^N \frac{\sigma_n^2}{k_n^4} \frac{\partial s_n(\theta)}{\partial \theta} \frac{\partial s_n(\theta)^T}{\partial \theta} \right)^{-1} \left( \sum_{i=1}^N \frac{1}{k_n^2} \frac{\partial s_n(\theta)}{\partial \theta} \frac{\partial s_n(\theta)^T}{\partial \theta} \right)^{-1}$$

As seen from equation (11), the performance of the *WLS* estimator depends on the noise sequence  $\sigma_1^2, \dots, \sigma_N^2$ , and on the predetermined weights,  $k_1, \dots, k_N$ . If the variance sequence happens to be equal to the weighting sequence squared, such that  $\sigma_n^2 = k_n^2$ , then

$$\text{Cov}(\hat{\theta}) = \left( \sum_{i=1}^N \frac{1}{\sigma_n^2} \frac{\partial s_n(\theta)}{\partial \theta} \frac{\partial s_n(\theta)^T}{\partial \theta} \right)^{-1} \quad (12)$$

which is exactly the *CRB* for this sequence. This result is in agreement with the fact that the weighted least squares estimator is the *ML* estimator when the noise is non stationary with a **known** noise variance sequence. It shows that, for a specific noise variance sequence, an  $M$  type estimator may be optimal and attain the *CRB*. Actually, in this problem we see that for any noise variance sequence, there exists an  $M$ -type estimator that attains the *CRB*. This estimator, however, cannot be constructed in advance since it requires prior knowledge of  $\sigma_n^2$ . That is, an  $M$ -type estimator is optimal for a specific sequence of the nuisance parameters but is suboptimal for other sequences. None of the  $M$ -type estimators are uniformly optimal for any noise sequences. Therefore, as first suggested by Huber [5], the choice of the estimating functions,  $\varphi_n$ , should be guided by robustness considerations.

## 5. SIMULATIONS - A SINUSOIDAL SIGNAL IN NOISE

A problem which appears in many applications is the estimation of parameters of a sinusoidal signal in noise. Consider the model:

$$y_n = A \sin(\omega n + \phi) + v_n \quad (13)$$

where  $A$ ,  $\phi$  and  $\omega$  are the amplitude, the phase and the frequency of the signal, respectively.  $v_n$  is a zero mean white, non stationary, Gaussian noise, i.e.,  $v_n \sim \mathcal{N}(0, \sigma_n^2)$ . In the following we describe simulation results for the  $M$  estimators of section 4.2.

We have considered the case where the frequency,  $\omega = 0.2\pi$ , is assumed known. the amplitude,  $A = 6$ , and the phase,  $\phi = 0$ , are to be estimated.

Computer simulations of three kinds of generalized  $M$ -type estimators were carried out: all of them are weighted least square estimators, with some weighting sequence  $k_n$ ,  $n = 1, 2, \dots, N$ . For the three cases, the choice of the weighting sequence is:

- $k_n = \sigma_n$ . This is the case where the weighting sequence happens to be the sequence of the (unknown) variance. Since theory shows that the performance of this estimate reaches the *CRB*, it is referred to as the "optimal" *WLS*.
- $k_n = 1$ . This classical LS estimator is the optimal estimator for the case of a stationary noise.
- Arbitrary, non constant  $k_n$ . This case is referred to as the suboptimal *WLS*.

The sequence of the values of the unknown noise variance,  $\sigma_n^2$ ,  $n = 1, 2, \dots, N$ , was generated randomly from Chi squared distribution with 10 degrees of freedom<sup>1</sup>. The weights of the suboptimal estimator were chosen as follows. For the first 100 samples, the weights were chosen such that  $k_n = \sigma_n + 1$ . For the rest of the samples, the weights were chosen such that  $k_n = |1 - \sigma_n|$ . It is anticipated that by this construction, the suboptimal estimator will be better then the least squares estimator in the first 100 samples (since it gives more weight to samples with small variance and less weight to samples with high variance). In the next 100 samples, the performance of the suboptimal *WLS* is expected to be ruined, since its weights are inversely related to the noise variance.

1500 Monte Carlo runs have been carried out. The (log) mean squared error of the estimates of the unknown amplitude and phase, respectively, as a function of the number of samples,  $N$ , is depicted in figures 1, 2.

The simulations results show:

1. The theoretical expression for the asymptotic performance of the estimates show good matching with the empirical results.
2. As expected, the performance of the optimal *WLS* is always the best, and coincides with the *CRB*.

<sup>1</sup>Our results, however, are not limited to the case where the sequence of the incidental parameters indeed comes from a distribution.

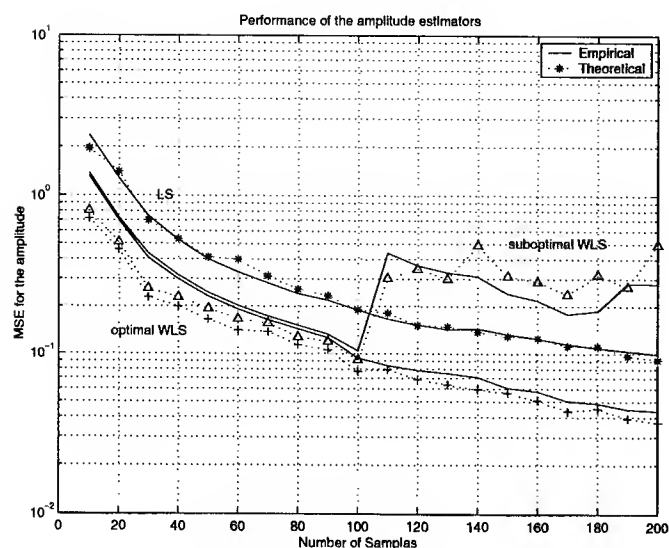


Figure 1: The MSE of the different estimates for the amplitude  $A$  versus the number of samples.

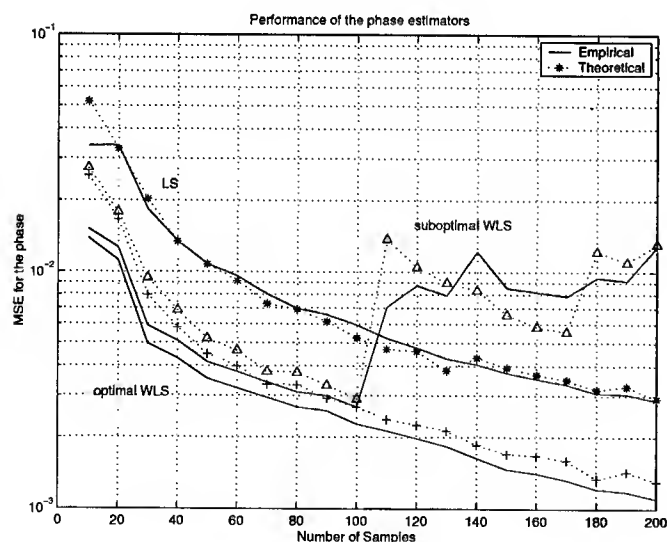


Figure 2: The MSE of the different estimates for the phase  $\phi$  versus the number of samples.

3. The performance of the LS and the (suboptimal) WLS can be better or worse than that of the LS, depending on the specific sequence of the values of the noise variance.

## 6. SUMMARY

In this paper we present a different approach to model non stationary processes. Instead of modeling the non stationarity in a parametric way with a small number of parameters, we propose to model it by unknown parameter set whose dimension increases with the number of observations. First, we study general flaws of such modeling. Then, focus is set on the problem of estimating the parameters of a deterministic signal in white additive non stationary Gaussian noise.

For the problem of interest, we propose the use of generalized  $M$  type estimates. The resulting estimates are analyzed and their asymptotic performance is evaluated analytically.

The WLS estimator is used as a demonstrating example for an  $M$  type estimator. This robust estimator, which is consistent under relatively mild conditions, is analyzed. Its asymptotic performance is evaluated analytically and is compared with simulation results through the problem of estimating the parameters of a sinusoid in non stationary, white, Gaussian noise.

## 7. REFERENCES

- [1] S. I. Amari and M. Kumon. Estimation in the presence of infinitely many nuisance parameters - geometry of estimating functions. *Annals of Statistics*, 16(3):1044-1068, 1988.
- [2] P. J. Bickel. On adaptive estimation. *Annals of Statistics*, 10:647-671, 1982.
- [3] J. Friedmann, E. Fishler, and H. Messer. Parameter estimation of a deterministic signal in non stationary, Gaussian noise. *submitted to IEEE Transactions on Signal Processing*.
- [4] R. Z. Hasminskii and I. A. Ibragimov. Efficient estimation in the presence of infinite dimensional incidental parameters. *Probability Theory and Mathematical Statistics. Lectures in Math.*, 1021:195-229, 1983.
- [5] P. J. Huber. *Robust Statistics*. John Wiley & Sons, 1981.
- [6] S. Mukhopadhyay and P. Sircar. Parametric modelling of non-stationary signals: A unified approach. *Signal Processing*, 60(2):135-152, July 1997.
- [7] J. Neyman and E. L. Scott. Consistent estimates based on partially consistent observations. *Econometrica*, 16(1):1-32, January 1948.



# CHANNEL ORDER AND RMS DELAY SPREAD ESTIMATION FOR AC POWER LINE COMMUNICATIONS \*

Hongbin Li<sup>1</sup> Zhaoqiang Bi<sup>2</sup> Duixian Liu<sup>3</sup> Jian Li<sup>2</sup> Petre Stoica<sup>4</sup>

<sup>1</sup>ECE Department, Stevens Institute of Technology, Hoboken, NJ 07030, USA (hli@stevens-tech.edu)

<sup>2</sup>ECE Department, University of Florida, Gainesville, FL 32611, USA

<sup>3</sup>Watson Research Center, IBM, Yorktown Heights, NY 10598, USA

<sup>4</sup>Systems & Control, Uppsala University, SE-751 03, Uppsala, Sweden

## Abstract

AC power lines have been considered as a convenient and low-cost medium for intra-building automation systems. In this paper, we investigate the problem of estimating the channel order and root mean squared (RMS) delay spread associated with the power lines, which are channel parameters that provide important information for determining the data transmission rate and designing appropriate equalization techniques for power line communications (PLC). We start by showing that the key to the RMS delay spread estimation problem is the determination of the channel order, i.e., the effective duration of the channel impulse response. We next discuss various ways to estimate the impulse response length from a noise-corrupted channel estimate. In particular, four different methods, namely a signal energy estimation (SEE) technique, a generalized Akaike information criterion (GAIC) based test, a generalized likelihood ratio test (GLRT), and a modified GLRT, are derived for determining the effective length of a signal contaminated by noise. These methods are compared with one another using both simulated and experimentally measured power line data. The experimental data was collected for power line characterization in frequencies between 1 and 60 MHz.

## 1 Introduction

AC power lines have been considered as a convenient carrier for communications in home and building automation systems [1]. However, power lines present a hostile environment for data transmission due to variable attenuation and impedance, impedance modulation, impulse noise, and continuous-wave jamming [1]. Recent applications of spread spectrum and forward error correction techniques to power line communications (PLC) have been quite successful in removing or alleviating the noise impediment [1]. Extensive char-

acterizations of power lines have also been reported [2]. However, these studies are mainly focused on frequencies up to 500 KHz. For video transmission and other similar applications using PLC, it is needed to characterize the PLC channel in the frequency range of 1 ~ 60 MHz.

In this paper, we are interested in the channel order and root mean squared (RMS) delay spread of power lines, which are important parameters that affect the data transmission rate over the channel. In practice, the PLC channel impulse response may be measured using some channel sounding techniques. The measured channel impulse response can be considered as the true impulse response contaminated by measurement noise which does not vanish with time. With this observation, one needs to determine from the noise-contaminated channel estimate the *effective* duration of the impulse response, outside which the measurements are primarily attributed to the noise. In this paper, we present a number of methods to solve the above effective signal length estimation problem. The performance of the various methods are compared using both simulated and experimentally measured data.

## 2 Problem Formulation

The impulse response of an AC power line,  $s(n)$ , is assumed to have a finite duration  $M$  [3] [4]. Given  $s(n)$  and the (effective) signal length  $M$ , the associated mean delay and RMS delay spread normalized with respect to the sampling interval can be determined as

$$\mu = \frac{\sum_{n=1}^M n s^2(n)}{\sum_{n=1}^M s^2(n)}, \quad (1)$$

and, respectively,

$$\sigma_{\text{RMS}} = \sqrt{\frac{\sum_{n=1}^M (n - \mu)^2 s^2(n)}{\sum_{n=1}^M s^2(n)}}. \quad (2)$$

In practice,  $s(n)$  can be estimated by measuring the frequency response of the power line channel using some channel sounding technique and applying the

\*This work was supported in part by the National Science Foundation Grant MIP-9457388, the Office of Naval Research Grant N00014-96-0817, and the Swedish Foundation for Strategic Research through a Senior Individual Grant.

inverse Fourier transform to the estimated frequency response [3]. The resulting signal  $x(n)$  can be considered as an estimate of  $s(n)$  contaminated by noise:

$$x(n) = s(n) + e(n), \quad n = 1, 2, \dots, N, \quad (3)$$

where  $e(n)$  denotes the estimation error which is modeled as a zero-mean white Gaussian noise with unknown variance  $\sigma_e^2$  and is assumed to be independent of  $s(n)$  [3] [4], and  $N$  is chosen such that  $N \gg M$ . The problem of interest here is to estimate the channel order  $M$  and the RMS delay spread  $\sigma_{\text{RMS}}$  from the measurements  $\{x(n)\}_{n=1}^N$ .

Supposing first that  $M$  is known, we can estimate  $\{s(n)\}_{n=1}^M$  by the maximum likelihood (ML) technique. Specifically, the negative log-likelihood function of  $\{x(n)\}$  is (see, e.g., [5])

$$V_M = \frac{N}{2} \ln \sigma_e^2 + \frac{1}{2\sigma_e^2} \left\{ \sum_{n=1}^M [x(n) - s(n)]^2 + \sum_{n=M+1}^N x^2(n) \right\} + \text{constant}. \quad (4)$$

The ML estimates of  $\sigma_e^2$  and  $s(n)$  are obtained by minimizing (4) with respect to the unknown parameters, which are given by  $\hat{\sigma}_e^2 = \frac{1}{N} \sum_{n=M+1}^N x^2(n)$  and  $\hat{s}(n) = x(n)$ ,  $n = 1, \dots, M$ , respectively. Thus,

$$\min_{\{s(n)\}, \sigma_e^2} V_M = \frac{N}{2} \ln \hat{\sigma}_e^2 + \text{constant}. \quad (5)$$

If  $M$  is known, we can replace  $s(n)$  in (1) and (2) by  $\hat{s}(n)$  to obtain an estimate of the RMS delay spread. The remaining question is how to estimate the signal length  $M$ , which is discussed next.

### 3 Signal Length Estimation

#### 3.1 SEE Based Test

Choose a sufficiently large  $L$  such that  $M \leq L \leq N$ . Let  $E_{LN}$  denote the total average energy of  $\{x(n)\}_{n=L}^N$ , i.e.,  $E_{LN} \triangleq \sum_{n=L}^N E[x^2(n)] = (N - L + 1)\sigma_e^2$ , where  $E[\cdot]$  denotes the expectation. The total noise average energy  $E_e$  is  $E_e \triangleq \sum_{n=1}^N E[e^2(n)] = \frac{N}{N-L+1} E_{LN}$ . Let  $E_x$  denote the total average energy of  $x(n)$ , i.e.,  $E_x \triangleq \sum_{n=1}^N E[x^2(n)]$ . The total deterministic signal energy is obtained as  $E_s \triangleq \sum_{n=1}^N s^2(n) = E_x - E_e$ . In practice,  $E_x$  and  $E_{LN}$  can be estimated as  $\hat{E}_x = \sum_{n=1}^N x^2(n)$  and  $\hat{E}_{LN} = \sum_{n=L}^N x^2(n)$ , respectively. It follows that an estimate of  $E_s$  is  $\hat{E}_s = \hat{E}_x - \frac{N}{N-L+1} \hat{E}_{LN}$ . The proposed SEE test consists of the following steps:

**Step 1:** Calculate  $\hat{E}_s$ .

**Step 2:** Set  $\check{M} = 1$  and  $\hat{E}'_s = 0$ .

**Step 3:** Compute  $\hat{E}'_s = \hat{E}'_s + x^2(\check{M}) - \frac{\hat{E}_e}{N}$ . Here, the updated  $\hat{E}'_s$  is the estimated total deterministic signal energy up to time index  $\check{M}$ .

**Step 4:** If  $\hat{E}'_s \geq \kappa \hat{E}_s$  or  $\check{M} = L$ , then the signal length estimate  $\hat{M}_{\text{SEE}}$  is equal to  $\check{M}$  and the test stops; otherwise, set  $\check{M} = \check{M} + 1$  and go to Step 3. Here,  $\kappa$  is a parameter of user choice, typically  $0.9 \leq \kappa \leq 0.99$ .

The SEE test is a method based on intuitive calculations of the signal and noise energies. It is simple but with a somewhat limited capability.

#### 3.2 GAIC Based Test

We describe here how to adopt the generalized Akaike information criterion (GAIC), original used for model structure selection in system identification [5], to determine the effective signal length  $M$ . The GAIC cost function has the form [5]:

$$\text{GAIC}_{\check{M}} = V_{\check{M}} + \gamma \ln(\ln N)(\check{M} + 1), \quad (6)$$

where  $V_{\check{M}}$  is defined in (5).  $\check{M}$  is assumed to be the signal length  $[(\check{M} + 1)$  is thus the total number of unknown parameters for the data model in (3)], and  $\gamma$  is a parameter of user choice. The proposed GAIC based test determines  $\hat{M}_{\text{GAIC}}$  by the following steps:

**Step 1:** Choose a sufficiently large  $L$  so that  $M \leq L \leq N$ .

**Step 2:** Calculate the cost function  $\text{GAIC}_{\check{M}}$  for  $\check{M} = 1, 2, \dots, L$ .

**Step 3:** The GAIC estimate of  $M$  is obtained as

$$\hat{M}_{\text{GAIC}} = \arg \min_{\check{M}} \text{GAIC}_{\check{M}}, \quad \check{M} = 1, 2, \dots, L. \quad (7)$$

*Remark 1:* As one may have noticed, using either the SEE or GAIC based test for determining  $M$  involves user parameters, viz  $\kappa$  in SEE and the  $\gamma$  in GAIC, which may affect the accuracy of the signal length estimate, but whose choice is not easy. Specifically, making a choice of these parameters to achieve a certain probability of detection (or missing) is not really possible. It would be desirable to derive methods that can somehow control the risk of making a wrong decision. Such methods should be of greater interest in real applications.

#### 3.3 GLRT

The generalized likelihood ratio for testing  $M = \check{M}$  against  $M = \check{M} + K$  (for some  $K \geq 1$ ) is given by [5]

$$\Lambda = N \ln \left[ \frac{\sum_{n=\check{M}+1}^N x^2(n)}{\sum_{n=\check{M}+K+1}^N x^2(n)} \right]. \quad (8)$$

For  $N \gg 1$  and under the hypothesis  $H_0: \check{M} \geq M$ , it can be shown that  $\Lambda$  is  $\chi^2$  distributed with  $K$  degrees

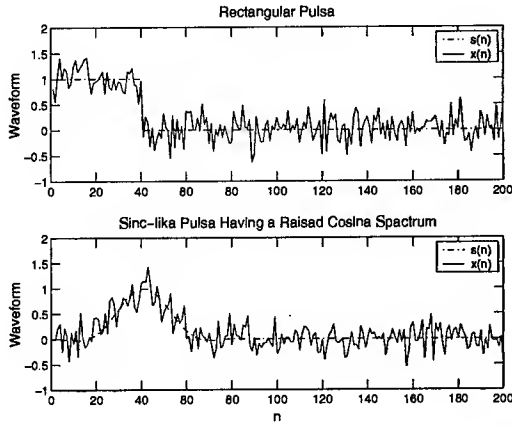


Figure 1: Tests signals used in the simulated examples.

of freedom, denoted by  $\Lambda \sim \chi^2(K)$ . To see this, we rewrite  $\Lambda$  as follows:

$$\Lambda = N \ln \left[ 1 + \frac{\sum_{n=\tilde{M}+1}^{\tilde{M}+K} x^2(n)}{\sum_{n=\tilde{M}+K+1}^N x^2(n)} \right] \quad (9)$$

$$N \gg \tilde{M}+K \quad N \frac{\sum_{n=\tilde{M}+1}^{\tilde{M}+K} x^2(n)}{\sum_{n=\tilde{M}+K+1}^N x^2(n)}.$$

Let  $\hat{\sigma}_e^2 = \frac{1}{N} \sum_{n=\tilde{M}+K+1}^N x^2(n)$ . Here,  $\hat{\sigma}_e^2$  is an estimate of  $\sigma_e^2$ . Note that for  $N \gg 1$ ,  $\hat{\sigma}_e^2 \approx \sigma_e^2$ . In view of this observation, we have (under  $H_0$ )

$$\Lambda \approx \frac{1}{\sigma_e^2} \sum_{n=\tilde{M}+1}^{\tilde{M}+K} x^2(n) \sim \chi^2(K). \quad (10)$$

The GLRT for determining  $\hat{M}_{\text{GLRT}}$  is summarized below:

**Step 1:** Choose a threshold  $\lambda$  from a table of the  $\chi^2$  distribution such that  $\Pr \{y \leq \lambda | y \sim \chi^2(K)\} = \alpha$ , where  $0.9 \leq \alpha \leq 0.99$  (see the discussions below).

**Step 2:** Set  $\tilde{M} = 1$ .

**Step 3:** Calculate  $\Lambda$  according to (8).

**Step 4:** If  $\Lambda \leq \lambda$  at  $\tilde{M}$  and also  $\Lambda \leq \lambda$  is true in more than 90% of the cases corresponding to  $\tilde{M} + 1, \tilde{M} + 2, \dots, L - K$ , then  $\hat{M}_{\text{GLRT}} = \tilde{M}$  and stop; otherwise, set  $\tilde{M} = \tilde{M} + 1$  and go to Step 3.

A brief explanation of Step 4 is as follows. It should be noted that  $K$  is a small integer, typically  $K \leq 10$  (see Section 3.4). However, a small  $K$  may be a bad choice for signals that are small over some intervals within the signal duration. When  $\tilde{M}$  happens to be in one of those intervals and also  $K$  is too small to include any significant signal energy in the denominator of (9), it is very likely that the inequality  $\Lambda \leq \lambda$  will be true. Hence, to find out the real signal boundary one has to

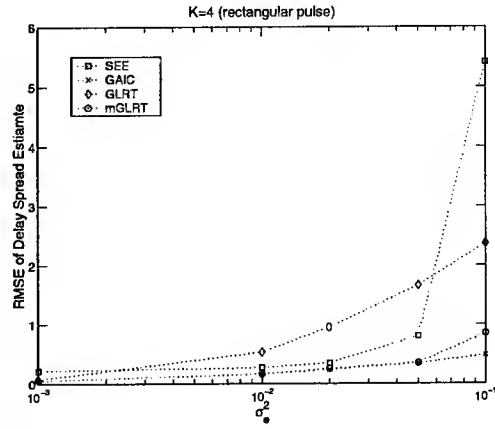


Figure 2: RMSE of  $\hat{\sigma}_{\text{RMS}}$  for the rectangular pulse versus  $\sigma_e^2$  when  $N = 450$ ,  $L = 200$ , and  $K = 4$ .

check the inequality  $\Lambda \leq \lambda$  not only at  $\tilde{M}$  but at the rest data samples as well. We shall keep in mind that even if the boundary sample has been hit,  $\Lambda \leq \lambda$  may not be true for all of the rest data samples due to the random nature of the noise. Nevertheless, the inequality should hold true for the majority (e.g., 90%) of the rest data samples beyond the signal boundary.

Observe that the risk of rejecting  $H_0$  when  $H_0$  holds (the probability of false alarm) equals  $1 - \alpha$ . In general, the risk of accepting  $H_0$  when it is not true cannot be determined for the statistics introduced previously unless one restricts considerably the class of alternative hypotheses against which  $H_0$  is tested. Thus, in applications the value of  $\alpha$  or, equivalently, the test threshold  $\lambda$  is chosen by considering only the probability of false alarm. Doing so, we shall keep in mind that as  $\alpha$  increases, the probability of false alarm decreases, but the other type of risk increases. Typically,  $\alpha$  is chosen between 0.9 and 0.99 [5].

*Remark 2:* It should be noted that the above GLRT is a valid test only when  $N \rightarrow \infty$ . Additionally,  $\hat{\sigma}_e^2$  is a poor estimate of  $\sigma_e^2$  if  $N$  is not large enough, particularly so if  $\tilde{M} + K + 1 < M$ . It would be of interest to modify the GLRT somehow such that the above problems are avoided. Such a modified GLRT indeed exists, as discussed next.

### 3.4 Modified GLRT

As mentioned before,  $\hat{\sigma}_e^2$  is not a good estimate of  $\sigma_e^2$ . A better estimate is  $\hat{\sigma}_e^2 = \frac{1}{N-L} \sum_{n=L+1}^N x^2(n)$ , where  $M \leq L \leq N$ . We now replace the  $\hat{\sigma}_e^2$  in (10) by the above  $\hat{\sigma}_e^2$  and define

$$\Delta \triangleq \frac{N-L}{K} \cdot \frac{\sum_{n=\tilde{M}+1}^{\tilde{M}+K} x^2(n)}{\sum_{n=L+1}^N x^2(n)} \triangleq \frac{N-L}{K} \cdot \frac{\rho_1}{\rho_2}. \quad (11)$$

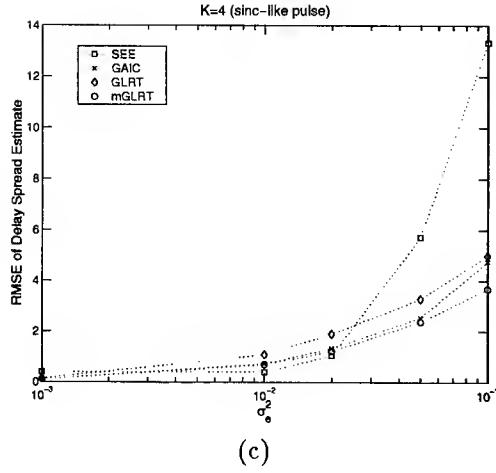


Figure 3: RMSE of  $\hat{\sigma}_{\text{RMS}}$  for the sinc-like pulse versus  $\sigma_e^2$  when  $N = 450$ ,  $L = 200$ , and  $K = 4$ .

Under the hypothesis  $H_0$  and if  $\check{M} + K \leq L$ ,  $\Delta$  is  $F$  distributed with  $K$  and  $N - L$  degrees of freedom [5], written as  $\Delta \sim F(K, N - L)$ . Observe that this holds in finite samples, whereas most other tests, including the original GLRT, require  $N \rightarrow \infty$ .

The choice of  $K$  should be made carefully. For  $\check{M} \geq M$ , this is perhaps not very important since any  $K \geq 1$  will lead to a similar performance. For  $\check{M} < M$ , however, the choice becomes more critical. To reduce the risk of underestimating  $M$ , a small  $K$  is recommended. To see this, let us assume that  $K$  is very large such that  $K \gg M$ . Then underestimating  $M$  by 1 or 2 will not increase  $\rho_1$  too much (particularly so if  $M$  is small), and hence the risk of underestimating  $M$  may be large. As a result, a smaller  $K$  in this case should be used. However, as mentioned in Section 3.3, a small  $K$  is a bad choice for signals that are small over certain intervals within the signal duration, and therefore a similar strategy may be used as adopted there. In particular, the modified GLRT determines  $\hat{M}_{\text{mGLRT}}$  in the following steps:

**Step 1:** Choose  $K \leq 10$  and a threshold  $\delta$  from a table of the  $F$  distribution so that  $\Pr\{z \leq \delta | z \sim F(K, N - L)\} = \alpha$ , where  $\alpha$  is between 0.9 and 0.99.

**Step 2:** Calculate  $\Delta$  for  $\check{M} = 1, 2, \dots, L - K$ .

**Step 3:** The signal length estimate  $\hat{M}_{\text{mGLRT}}$  is the smallest  $\check{M}$  at which  $\Delta \leq \delta$  is true and for which the inequality is also true in more than  $90\alpha\%$  of the cases corresponding to  $\check{M} + 1, \check{M} + 2, \dots, L - K$ .

## 4 Numerical Results

In the following, we use  $\kappa = 0.96$  for the SEE test,  $\gamma = 2$  for GAIC,  $K = 4$  and  $\alpha = 0.99$  for GLRT and the modified GLRT (referred to as mGLRT henceforth).

### 4.1 Simulated Example

The simulated data consists of a pulse having a certain shape corrupted by a zero-mean white Gaussian noise with variance  $\sigma_e^2$ . We consider both a rectangular pulse ( $M = 40$ ) and a sinc-like pulse having a raised cosine spectrum. The roll-off factor for the latter is 1. The sinc-like pulse is shifted and truncated to have a duration of 80 samples. Figure 1 shows a realization of the test data corresponding to the two different pulses when  $\sigma_e^2 = 0.05$ , where dashdot lines represent noise-free signals and solid lines denote noise-contaminated signals, respectively. The results shown below are obtained using 200 Monte-Carlo trials. For each individual trial, a total number of  $N = 450$  samples are used and  $L = 200$ . In this example, we investigate the effect of the noise variance on the performance of the proposed tests. The root mean squared error (RMSE) of  $\hat{\sigma}_{\text{RMS}}$  are shown in Figure 2 for the rectangular pulse and Figure 3 for the sinc-like pulse. The results show that in general mGLRT and GAIC perform better than the other two methods. The poor performance of GLRT is due to that a number of *outliers*, i.e., the signal length estimates are equal to  $L$ , are obtained by GLRT.

### 4.2 Experimental Example

We now consider an experimental example. We first briefly describe the PLC channel sounding system used to obtain the measurement data. For more details of the system and measurement process, we refer the interested readers to [3]. Figure 4 shows a block diagram that uses impulse channel sounding to measure the impulse response of the AC power line channel. The coupler box plugging into the AC wall outlet (the top path in Figure 4) behaves like a highpass filter, with the 3 dB cutoff at 1 MHz. The probing signal passes through the coupler and the AC power line network and exits through a similar coupler plugged in a different outlet. A direct coupler to coupler connection is used to calibrate the test setup (the bottom path in Figure 4). A low-noise amplifier (LNA) with at least 54 dB gain is used in front of the digital storage oscilloscope (DSO) to reduce the noise figure and increase the sensitivity of the system. The LNA has a built-in lowpass filter with the 3 dB cutoff frequency at 70 MHz in the front stage. Additionally, a high-precision adjustable (0-80 dB) attenuator is placed after the receiving coupler, making it possible to center the dynamic range of the LNA/DSO combination for the signal level of each outlet pair. This allows the system to capture noise spikes and temporal noise fluctuations. The DSO has a bandwidth of 500 MHz, implying a high resolution, and the capability for long time captures.

The probing impulse used is a specially truncated

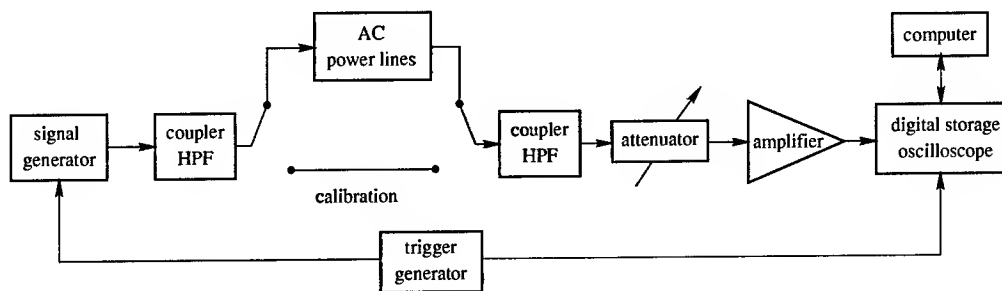


Figure 4: Power line channel measurement system.

sinc waveform, with a duration of 17 ns and a flat frequency characteristics from 0.85 to 63.6 MHz. The highpass characteristics of the couplers and the low-pass filter in the LNA limit the receiving sensitivity of the system to the 1 to 60 MHz frequency band. The sampling frequency is 1 GHz and the total number of data samples is  $N = 20000$ . The measurements were performed at two residential houses by averaging over 100 to 1000 scope sweeps depending on the noise situation. Figure 5 shows the impulse response of a specific power line channel corresponding to the frequency band 1 ~ 60 MHz. For channel order and RMS delay spread estimation, we choose  $L = N/2$ . The effective signal length estimates obtained by the four tests under study are also shown in the figure. We notice that GLRT fails again since the GLRT estimate is an outlier, with a value equal to  $L$ . It is also seen that SEE obviously underestimates the effective signal length. On the other hand, the estimates given by GAIC and mGLRT appear to be more accurate. After obtaining the effective signal length estimate, we can use (1) and (2) to calculate the mean delay and RMS delay spread. Specifically, the RMS delay spread estimates for the 1 ~ 60 MHz frequency band obtained by SEE, GAIC, and mGLRT are 0.19, 0.27, and 0.28  $\mu$ s, respectively. With no equalization, the maximum transmission rate is inversely proportional to the RMS delay spread, i.e., Maximum Transmission Rate  $\approx \frac{1}{2\sigma_{\text{RMS}}}$ . It follows that the maximum data transmission rate is approximately 2.63 Mbps. The above calculation is somewhat optimistic since other factors, such as attenuation and noise characteristics of the PLC channel, which are important in determining the transmission rate, were not counted. Additionally, the impulse responses were obtained using one specific set of measurements. It is our experience that the RMS delay spread could vary significantly depending on the loads and environment of the power lines networks.

## References

[1] D. Radford, "Spread-spectrum data leap through

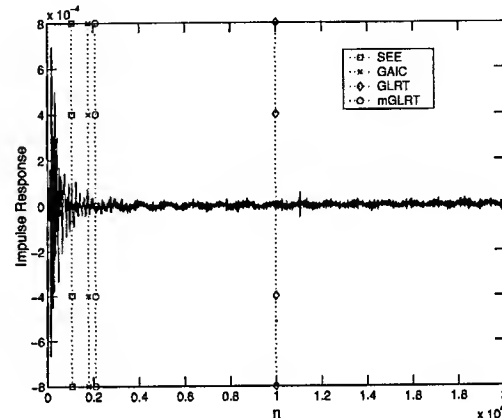


Figure 5: Impulse response of the power line channel (1 ~ 60 MHz) and the corresponding effective signal length estimates.

ac power wiring," *IEEE Spectrum*, pp. 48-53, November 1996.

- [2] M. H. L. Chan and R. W. Donaldson, "Attenuation of communication signals on residential and commercial intrabuilding power-distribution circuits," *IEEE Transactions on Electromagnetic Compatibility*, vol. 28, pp. 220-230, November 1986.
- [3] D. Liu, E. Flint, B. Gaucher, and Y. Kwark, "Wide band AC power line characterization," *IEEE Transactions on Consumer Electronics*, November 1999.
- [4] D. Liu, E. Flint, B. Gaucher, and Y. Kwark, "High-speed communications system performance using the home AC power line," Technical Report No. RC21604(97420), IBM T. J. Watson Research Center, November 1999.
- [5] T. Söderström and P. Stoica, *System Identification*. London, UK: Prentice Hall International, 1989.

# Taylor Series Adaptive Processing

Daniel J. Rabideau

Massachusetts Institute of Technology  
Lincoln Laboratory  
Lexington, MA 02420, USA

## ABSTRACT

Many signal processing applications require estimating and tracking a quantity that is inherently nonstationary. Such quantities may be matrices (e.g., a covariance matrix or an image), vectors (e.g., a weight vector or an eigenvector), or scalars. This paper considers the use of Taylor series expansions to enhance tracking. The potential benefits of this approach include: (1) a reduction in computational burden, (2) a reduction in required memory size and/or communication bandwidth (via an implicit compression of the quantity of interest), (3) interpolation through "gaps" in the available data, and (4) increased fidelity due to the explicit incorporation of "nonstationarity" into the model. Sensor array processing examples are used to illustrate the approach.

## 1. INTRODUCTION

Many signal processing applications require estimating and tracking a quantity that is inherently nonstationary. Typically, sensor data is collected and used to form an initial estimate of the quantity. Later, more sensor data is collected and a new estimate is formed. The process is repeated, thereby "tracking" the quantity of interest.

There are several problems with this "estimation – then – re-estimation" approach. First, the estimation procedure is often computationally intensive. Computing estimates may take a very long time and may be difficult to implement in real-time. Once an initial estimate is found, it is thus desirable to have an efficient method for tracking the quantity (e.g., via recursive least squares approaches or "fast" subspace tracking).

Second, the estimation procedure is typically formulated on the assumption that the sensor data is stationary within the "training" interval (i.e., the interval used for creating the estimate). Often this is only approximately true, and thus may degrade performance.

Third, there may be gaps in the available data that make tracking difficult or impossible.

Finally, the tracked quantity must often be stored. Sometimes, it must also be transmitted over a wireless link (e.g., for remote processing). In this case, frequent re-estimation can drive-up the storage requirements and/or link bandwidth. Compression of the quantity is thus a desirable alternative;

however, this is typically only done after the estimates are formed (not as an integral part of the tracking algorithm).

As an illustration of a representative application, consider sensor array processing (e.g., as applied to radar, sonar, or wireless communications). Typically, sensor array snapshots are used to estimate a covariance matrix,  $\mathbf{R}$ , its inverse,  $\mathbf{R}^{-1}$ , its principal subspace,  $\mathbf{V}_s$ , an image, a spectrum, or other quantities. Such quantities are used in adaptive beamforming, adaptive Doppler filtering, STAP, MUSIC, ESPRIT, MVDR, and other super-resolution methods.

When the covariance matrix is desired, its maximum likelihood estimate (presuming stationary interference) is typically used:

$$\hat{\mathbf{R}} = \frac{1}{K} \sum_{k=1}^K \mathbf{x}(k) \mathbf{x}^H(k) \quad (1)$$

where  $\mathbf{x}(k)$  is the  $k^{\text{th}}$  array snapshot. When principal components are needed, the eigenvalue decomposition of  $\hat{\mathbf{R}}$  is first calculated,

$$\hat{\mathbf{R}} = \hat{\mathbf{V}} \hat{\mathbf{D}} \hat{\mathbf{V}}^H. \quad (2)$$

The principal subspace,  $\hat{\mathbf{V}}_s$ , can then be found by extracting the columns of  $\hat{\mathbf{V}}$  associated with the  $s$  largest eigenvalues (i.e., the  $s$  diagonal elements of  $\hat{\mathbf{D}}$  above the noise floor).

In nonstationary environments, such quantities are functions of time, i.e.,  $\mathbf{R}(t)$ ,  $\mathbf{R}^{-1}(t)$ , and  $\mathbf{V}_s(t)$  respectively. Estimates of these quantities are typically made via the standard estimators (i.e., (1) and (2)) applied to a training interval consisting of nearby snapshots, e.g. (for odd  $K$ ):

$$\hat{\mathbf{R}}(t) = \frac{1}{K} \sum_{k=t-\frac{K-1}{2}}^{t+\frac{K-1}{2}} \mathbf{x}(k) \mathbf{x}^H(k) \quad (3)$$

$$\hat{\mathbf{R}}(t) = \hat{\mathbf{V}}(t) \hat{\mathbf{D}}(t) \hat{\mathbf{V}}^H(t) \quad (4)$$

As  $t$  varies, tracking is achieved by recalculating these estimators. Some efficient techniques exist for updating these quantities as long as there is a large degree of overlap between successive training intervals. In practice, this may not be the case. Furthermore, the data within the training interval may be nonstationary.

In Section 2, we consider nonstationary models for the *estimated* quantities. We investigate ways that Taylor series expansions may be employed in the tracking of these estimated quantities. In Section 3, we consider nonstationary models for the *underlying* quantities. We investigate ways that Taylor series expansions may be used to estimate these underlying quantities.

This work was sponsored by U.S. Navy under Air Force Contract F19628-95-C-0002. Opinions, interpretations, conclusions and recommendations are those of the author and are not necessarily endorsed by the United States Air Force.

Our approach explicitly models the nonstationarity, and can result in better performance than standard estimators. Section 4 contains a summary.

## 2. TRACKING ESTIMATED QUANTITIES

As an introduction, let us consider how Taylor series expansions can be applied to the problem of tracking *estimated* quantities (note: the more interesting problem of tracking *underlying* quantities is considered in the next section). Suppose  $\hat{\mathbf{A}}(t)$  is an estimate of a quantity of interest. Without loss of generality, let the dimensions of  $\hat{\mathbf{A}}(t)$  be  $N \times P$ . The Taylor series expansion of  $\hat{\mathbf{A}}(t)$  about  $t = 0$  is:

$$\hat{\mathbf{A}}(t) = \hat{\mathbf{A}}(0) + t\dot{\hat{\mathbf{A}}}(0) + \dots + \frac{t^n}{n!} \hat{\mathbf{A}}^{(n)}(0) + H.O.T. \quad (5)$$

where  $\dot{\hat{\mathbf{A}}}(0)$  denotes the first derivative,  $\hat{\mathbf{A}}^{(n)}(0)$  denotes the  $n^{\text{th}}$  derivative, and "H.O.T." denotes higher order terms of the series.

Suppose estimates of this quantity are made at  $t_1, t_2, \dots, t_E$ . A system of  $n^{\text{th}}$  order Taylor Series approximation equations can then be written:

$$(\mathbf{T} \otimes \mathbf{I}_{N \times N}) \begin{bmatrix} \hat{\mathbf{A}}(0) \\ \dot{\hat{\mathbf{A}}}(0) \\ \vdots \\ \hat{\mathbf{A}}^{(n)}(0) \end{bmatrix} = \begin{bmatrix} \hat{\mathbf{A}}(t_1) \\ \hat{\mathbf{A}}(t_2) \\ \vdots \\ \hat{\mathbf{A}}(t_E) \end{bmatrix} \quad (6)$$

where  $\otimes$  denotes the kronecker product, and

$$\mathbf{T} = \begin{bmatrix} 1 & t_1 & \dots & \frac{t_1^n}{n!} \\ 1 & t_2 & \dots & \frac{t_2^n}{n!} \\ \vdots & \vdots & \dots & \vdots \\ 1 & t_E & \dots & \frac{t_E^n}{n!} \end{bmatrix} \quad (7)$$

Solving this system of equations will provide estimates of  $\hat{\mathbf{A}}(0), \dot{\hat{\mathbf{A}}}(0), \dots, \hat{\mathbf{A}}^{(n)}(0)$ . Observe that a solution is given by:

$$\begin{bmatrix} \hat{\mathbf{A}}(0) \\ \dot{\hat{\mathbf{A}}}(0) \\ \vdots \\ \hat{\mathbf{A}}^{(n)}(0) \end{bmatrix} = ((\mathbf{T}^H \mathbf{T})^{-1} \mathbf{T}^H \otimes \mathbf{I}_{N \times N}) \begin{bmatrix} \hat{\mathbf{A}}(t_1) \\ \hat{\mathbf{A}}(t_2) \\ \vdots \\ \hat{\mathbf{A}}(t_E) \end{bmatrix} \quad (8)$$

In the event that the system of (6) is overdetermined or underdetermined, (8) provides the least squares solution. Also note that the psuedo-inverse in (8) can often be pre-computed, or written in closed form.<sup>†</sup>

<sup>†</sup> In some cases, (8) can then be re-factored so as to combine the two terms on the right directly, rather than separately forming the  $\hat{\mathbf{A}}(t_i)$ 's and  $(\mathbf{T}^H \mathbf{T})^{-1} \mathbf{T}^H \otimes \mathbf{I}_{N \times N}$ .

### Examples:

Next, let us illustrate the usefulness of (8) for tracking estimated quantities. Suppose we have a uniform linear array of 20 half-wavelength spaced isotropic elements. Two nonstationary signal sources are present; their initial angles are  $6.418^\circ$  and  $-18.998^\circ$  relative to array broadside (this a separation of 10 beamwidths—where a "beamwidth" is defined here as the angle between the peak and 3 dB points of the beam pattern). They each have an SNR of 30 dB. A total of 99 samples are collected (i.e.,  $t$  varies from -49 to 49); each source moves 5 beamwidths during this time (thus, the first source passes through  $0^\circ$  at  $t = 0$ ).

Suppose we wish to track the spectrum (a.k.a. periodogram) vs. time,

$$P(\theta, t) \equiv \mathbf{v}(\theta)^H \mathbf{R}(t) \mathbf{v}(\theta) \quad (9)$$

where  $\mathbf{v}(\theta)$  is the array response vector for a source at angle  $\theta$ . Figure 1a shows the true instantaneous periodogram calculated from (9). In practice, the true instantaneous covariance matrix,  $\mathbf{R}(t)$ , would be unknown and must be estimated. Thus, we would actually compute:

$$\hat{P}(\theta, t) \equiv \mathbf{v}(\theta)^H \hat{\mathbf{R}}(t) \mathbf{v}(\theta) \quad (10)$$

where  $\hat{\mathbf{R}}(t)$  is an estimate of the covariance matrix as in (3). Since the sources are moving, the sample averaging in (3) will introduce an error. Thus, we should like to use a small  $K$  (but if  $K$  is too small, the spectrum will be noisy and sources may be incorrectly estimated). Figure 1b shows the estimated periodogram calculated from (10) and (3) with  $K = 1$ . As expected, this estimate is very noisy and its peaks are frequently in the wrong places. To address this, Figure 1c shows the periodogram when  $K = 21$ . Here, the covariance matrix and periodogram are re-estimated at each time,  $t$  (except near the ends, where the desired training interval falls outside the available data).

Finally, Figure 1d shows the estimated periodogram that results when Taylor-series approximations are used. Here, five sample covariances were formed using (3) (with  $K = 21$  and  $t = 11, 30, 49, 69$  and  $87$ , respectively). Then, we used (8) and (5) (with  $\hat{\mathbf{A}}(t_i)$  replaced by  $\hat{\mathbf{R}}(t_i)$ ,  $E = 5$ , and  $n = 4$ ) to synthesize all of the  $\hat{\mathbf{R}}(t)$  values used in (10).

Observe that the Taylor-series method performed very similar to the "estimate - then - re-estimate" approach shown in Figure 1c (even though we only formed 5 sample covariances from the data!). In fact, the measured root-mean-squared-error (RMSE) between the peaks<sup>‡</sup> of Figure 1c and Figure 1a (excluding the invalid  $t$ 's near both ends) was  $0.21^\circ$ . By comparison, the measured RMSE between the peaks of Figure 1d and Figure 1a was  $0.18^\circ$  within this same time interval. (Outside this interval, the Taylor series approach appears to also provide a reasonable extrapolation!).

<sup>‡</sup> The search for peaks was conducted on a  $0.1^\circ$  grid.

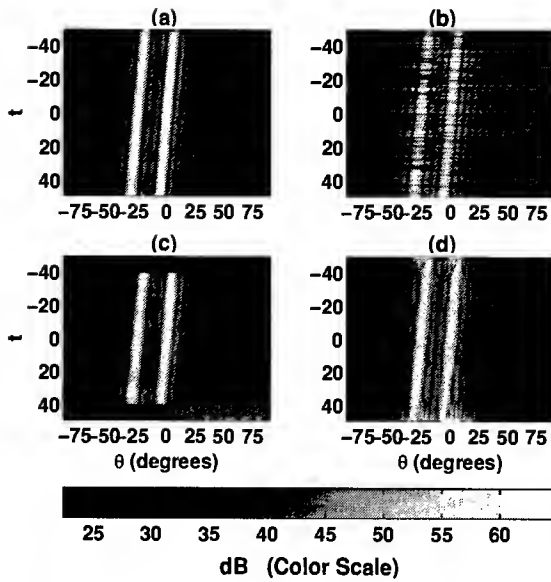


Figure 1. Periodograms. (a)  $P(\theta, t)$ . (b)  $\hat{P}(\theta, t)$  using (3) and  $K = 1$ . (c)  $\hat{P}(\theta, t)$  using (3) and  $K = 21$ . (d)  $\hat{P}(\theta, t)$  using Taylor-series approximations.

It is perhaps worth noting that our Taylor-based estimates may be used directly in their "series form." For example, the periodogram above may be written:

$$\hat{P}(\theta, t) \approx \mathbf{v}(\theta)^H \hat{\mathbf{R}}(0) \mathbf{v}(\theta) + \dots + \frac{t^n}{n!} \mathbf{v}(\theta)^H \hat{\mathbf{R}}^{(n)}(0) \mathbf{v}(\theta). \quad (11)$$

This quantity is efficiently updated at multiple  $t$ 's by calculating the quadratic terms only once, and then forming different weighted combinations at each different  $t$ <sup>§</sup>. Moreover, storage and/or transmission (over a wireless link) of the result is simplified by maintaining the series form. Of course, the same ideas apply to other quantities as well (e.g., the computation of adaptive beamformer weights via Taylor series expansion of  $\hat{\mathbf{R}}(t)^{-1}$ ).

As a second illustration, consider Figure 2. Here, we illustrate the effect of a gap in the sensor data. Snapshots at  $t = -10, \dots, 0$  are removed. Sample covariance estimates (with  $K = 21$ ) are formed at  $t = -39 \dots -21, 11 \dots 39$ . Then, we solve for the Taylor series terms and use them to synthesize covariance estimates at all  $t$  (including the gap region). Comparing Figure 2 and Figure 1c, performance is obviously pretty good.

As a final illustration, we applied Taylor series to the modeling of the estimated MVDR spectrum,

$$\hat{M}(\theta, t) \equiv \frac{1}{\mathbf{v}(\theta)^H \hat{\mathbf{R}}(t)^{-1} \mathbf{v}(\theta)} \quad (12)$$

Figure 3a show the spectrum using the usual sample covariance estimates (with  $K = 21$ , and  $\hat{M}(\theta, t)$  recomputed at each  $t$ ). Figure 3b shows the spectrum resulting from Taylor expansion of

$\hat{M}(\theta, t)$ . Here, we used  $E = 40$  (i.e., we skipped every other  $t$ ) and  $n = 7$ . Clearly, the method provides an adequate interpolation at the missing  $t$ 's, though the sidelobes are noticeably higher.

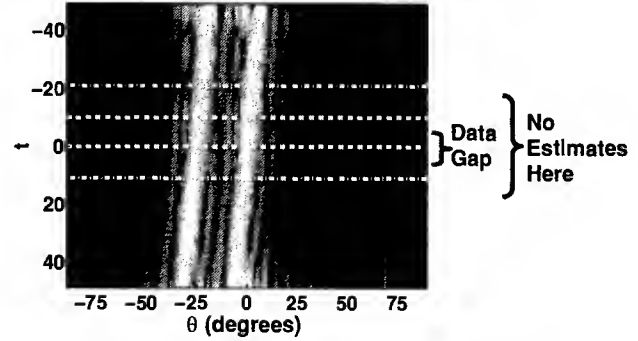


Figure 2. Taylor Series approximation to periodogram, with gap in available data.

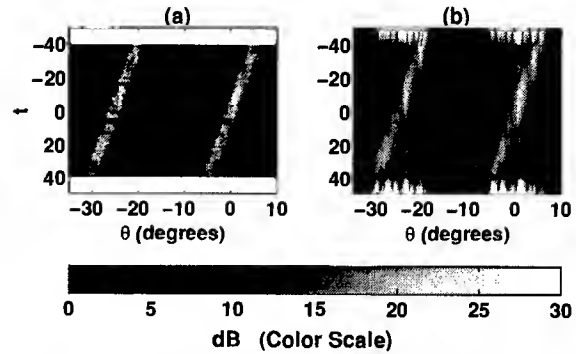


Figure 3. MVDR Spectrum (a) Estimated every  $t$  via (3) with  $K = 21$  (b) Estimated by Taylor series with many small gaps between measurements.

### 3. TRACKING UNDERLYING QUANTITIES

In Section 2, we used Taylor series expansions to model estimated quantities, e.g.,  $\hat{\mathbf{A}}(t)$ . Our model was then used to interpolate, extrapolate and smooth our estimates (while also potentially reducing computation, data storage and/or communication bandwidth).

To summarize Section 2, our basic procedure was to:

- (1) Initially estimate  $\hat{\mathbf{A}}(t_i)$  at several  $t_i$ ,
- (2) Solve for the terms in the Taylor series expansion of  $\hat{\mathbf{A}}(t)$
- (3) Use these terms to interpolate, smooth, etc.

Looking back at Section 2, we may now ask the question: "Does this procedure make logical sense?" Conceptually, the procedure correctly regards  $\hat{\mathbf{A}}(t)$  as being "nonstationary" during step 2 -- the computation of the Taylor series expansion terms. However, step 1 leads to a bit of a dilemma. In the scenarios that interest

<sup>§</sup> Of course, we could just expand  $\hat{P}(\theta, t)$  directly, instead of  $\hat{\mathbf{R}}(t)$ .



us most (and probably many other scenarios as well), the available procedures for initially estimating  $\hat{\mathbf{A}}(t_i)$  are implicitly based upon an assumption that the training data is stationary. Consider, for example, the sample covariance estimate,  $\hat{\mathbf{R}}(t)$ , in (3). In the sensor array processing community, this estimator is very widely used. It has the property of being a maximum likelihood estimator when the observations,  $\mathbf{x}(t)$ , are stationary. However, it is also frequently used in situations where the data is *not* stationary. Could we do better?

### 3.1 Tracking $\mathbf{R}(t)$ or $P(\theta, t)$

To answer this question, let us begin by expanding the "underlying" (i.e., true instantaneous) covariance matrix:

$$\mathbf{R}(t) = \mathbf{R}(0) + t\dot{\mathbf{R}}(0) + \dots + \frac{t^n}{n!} \mathbf{R}^{(n)}(0) + H.O.T. \quad (13)$$

Next, use this to expand the expected value of the sample covariance estimator in (3),

$$\begin{aligned} E\{\hat{\mathbf{R}}(t)\} &= \frac{1}{K} \sum_{k=1}^{K-1} \mathbf{R}(k) \\ &= \mathbf{R}(0) + t\dot{\mathbf{R}}(0) + \dots + \frac{1}{K} \sum_{k=1}^{K-1} \frac{k^n}{n!} \mathbf{R}^{(n)}(0) + H.O.T. \end{aligned} \quad (14)$$

From (14), we can create a system of  $n^{\text{th}}$  order Taylor series approximation equations:

$$\begin{pmatrix} \tilde{\mathbf{T}}^H \otimes \mathbf{I}_{N \times N} \end{pmatrix} \begin{bmatrix} \mathbf{R}(0) \\ \dot{\mathbf{R}}(0) \\ \vdots \\ \mathbf{R}^{(n)}(0) \end{bmatrix} = \begin{bmatrix} E\{\hat{\mathbf{R}}(t_1)\} \\ E\{\hat{\mathbf{R}}(t_2)\} \\ \vdots \\ E\{\hat{\mathbf{R}}(t_E)\} \end{bmatrix} \quad (15)$$

where

$$\tilde{\mathbf{T}} \equiv \begin{bmatrix} 1 & t_1 & \dots & \frac{1}{K} \sum_{k=t_1}^{t_1 + \frac{K-1}{2}} \frac{k^n}{n!} \\ 1 & t_2 & \dots & \frac{1}{K} \sum_{k=t_2}^{t_2 + \frac{K-1}{2}} \frac{k^n}{n!} \\ \vdots & \vdots & \dots & \vdots \\ 1 & t_E & \dots & \frac{1}{K} \sum_{k=t_E}^{t_E + \frac{K-1}{2}} \frac{k^n}{n!} \end{bmatrix} \quad (16)$$

The least squares solution to (15) is then given by:

$$\begin{bmatrix} \mathbf{R}(0) \\ \dot{\mathbf{R}}(0) \\ \vdots \\ \mathbf{R}^{(n)}(0) \end{bmatrix} = \left( (\tilde{\mathbf{T}}^H \tilde{\mathbf{T}})^{-1} \tilde{\mathbf{T}}^H \otimes \mathbf{I}_{N \times N} \right) \begin{bmatrix} E\{\hat{\mathbf{R}}(t_1)\} \\ E\{\hat{\mathbf{R}}(t_2)\} \\ \vdots \\ E\{\hat{\mathbf{R}}(t_E)\} \end{bmatrix} \quad (17)$$

Hence if we could solve equation (17), we could then use (13) (with terms above order  $n$  truncated) to estimate the underlying function  $\mathbf{R}(t)$ . In practice, the expected values on the right side of (17) are unknown. Instead, we will replace these expectations by single realizations. That is, we'll solve:

$$\begin{bmatrix} \mathbf{R}(0) \\ \dot{\mathbf{R}}(0) \\ \vdots \\ \mathbf{R}^{(n)}(0) \end{bmatrix} = \left( (\tilde{\mathbf{T}}^H \tilde{\mathbf{T}})^{-1} \tilde{\mathbf{T}}^H \otimes \mathbf{I}_{N \times N} \right) \begin{bmatrix} \hat{\mathbf{R}}(t_1) \\ \hat{\mathbf{R}}(t_2) \\ \vdots \\ \hat{\mathbf{R}}(t_E) \end{bmatrix} \quad (18)$$

After solving for  $\mathbf{R}(0)$ ,  $\dot{\mathbf{R}}(0)$ , ...,  $\mathbf{R}^{(n)}(0)$  in (18) we'll use them in (13) to synthesize interpolated, smoothed, or extrapolated values for  $\mathbf{R}(t)$ .

**Example:**

Let us next illustrate the benefit of tracking the underlying  $\mathbf{R}(t)$  via (18) and (13). Assume the same array scenario as in Section 2. A single source is present and moving in a nonlinear fashion. Its true angle is:

$$\theta(t) = (t-1)^2 \varepsilon \quad (19)$$

where  $\varepsilon = (2/99)^\circ$ . As earlier, a total of 99 snapshots are available.

Figure 4a shows the true instantaneous periodogram calculated from (9). In contrast, Figure 4b shows the estimated periodogram calculated from (10) and (3) with  $K = 21$ . Observe how the sample averaging now causes large errors. We wish to compensate these errors in the neighborhood of  $t = 0$ . Figure 4c attempts this by using the method of (18) and (13). Here,  $K = 21$ ,  $E = 79$ ,  $n = 9$ . The RMSE's were evaluated for the methods of Figure 4b and Figure 4c, and averaged over the region  $t = -10 \dots 10$ . A total of 100 independent trials were performed and the average RMSE was found to be  $0.6416^\circ$  for the method of Figure 4b, and  $0.4701^\circ$  for the method of Figure 4c. Hence, the covariance estimation method of (18), (13) has resulted in a 27% reduction in the angle estimation error. Incidentally, Figure 4d shows the method when  $K = 1$  (in this limit, equations (8) and (18) are the same).

### 3.2 Tracking $\mathbf{V}_s(t)$

As a final case study, let's consider  $\mathbf{V}_s(t)$ . Using the method of Section 2, we can create Taylor series approximations to  $\hat{\mathbf{V}}_s(t)$ . However, we would prefer to track the underlying  $\mathbf{V}_s(t)$  and thus compensate for the averaging used in forming the estimate  $\hat{\mathbf{R}}(t)$ . How do we do this?

Let us begin by expanding  $\mathbf{V}_s(t)$  in a Taylor series:

$$\mathbf{V}_s(t) = \mathbf{V}_s(0) + t\dot{\mathbf{V}}_s(0) + \dots + \frac{t^n}{n!} \mathbf{V}_s^{(n)}(0) + H.O.T. \quad (20)$$

\*\* Note: the basis vectors synthesized in this manner will not generally be orthonormal.

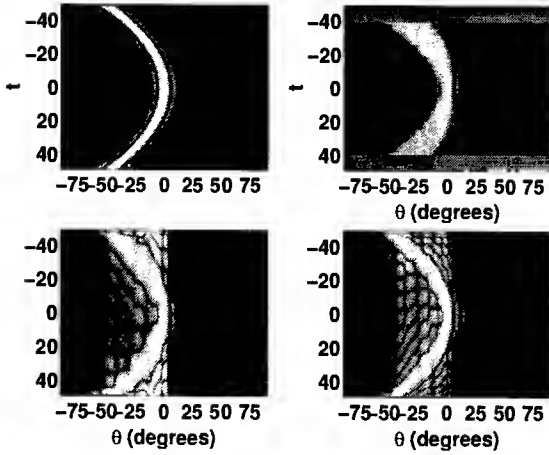


Figure 4. Periodograms. (a)  $P(\theta, t)$ . (b)  $\hat{P}(\theta, t)$  using (3) and  $K = 21$ . (c) Taylor-series approximation that compensates for nonstationarity near  $t = 0$ . (d) Taylor-series approximation with  $K = 1$ ,  $n = 14$ . [Note: 30 dB color scale similar to Fig. 3]

Column  $i$  holds the  $i^{\text{th}}$  unit-normalized basis vector:

$$\mathbf{v}_{s,i}(t) = \mathbf{v}_{s,i}(0) + t\dot{\mathbf{v}}_{s,i}(0) + \dots + \frac{t^n}{n!} \mathbf{v}_{s,i}^{(n)}(0) + H.O.T. \quad (21)$$

By definition,  $\mathbf{v}_{s,i}(t)$  maximizes

$$\mathbf{v}_{s,i}^H(t) E\{\mathbf{x}(t)\mathbf{x}^H(t)\} \mathbf{v}_{s,i}(t) \quad (22)$$

over the set of all vectors orthogonal to  $\mathbf{v}_{s,u}(t)$ ,  $u \in \{1, \dots, i-1\}$ .

Thus, we wish to approximately maximize

$$\tilde{\mathbf{v}}_{s,i}^H E\{\mathbf{x}(t)\mathbf{x}^H(t)\} \tilde{\mathbf{v}}_{s,i} \quad (23)$$

where  $\tilde{\mathbf{v}}_{s,i} = \mathbf{v}_{s,i}(0) + \dots + \frac{t^n}{n!} \mathbf{v}_{s,i}^{(n)}(0)$ . Equation (23) can be rewritten as:

$$\mathbf{u}_{s,i}^H E\{\mathbf{y}(t)\mathbf{y}^H(t)\} \mathbf{u}_{s,i} \quad (24)$$

where  $\mathbf{u}_{s,i}^H = \left[ \mathbf{v}_{s,i}^H(0) \quad t\dot{\mathbf{v}}_{s,i}^H(0) \quad \dots \quad \frac{t^n}{n!} \mathbf{v}_{s,i}^{(n)H}(0) \right]$  and

$\mathbf{y}^H(t) = [\mathbf{x}^H(t) \quad \mathbf{x}^H(t) \quad \dots \quad \mathbf{x}^H(t)]$ . This, in turn, is equal to:

$$\mathbf{t}_{s,i}^H E\{\mathbf{z}(t)\mathbf{z}^H(t)\} \mathbf{t}_{s,i} \quad (25)$$

where  $\mathbf{t}_{s,i}^H = \left[ \mathbf{v}_{s,i}^H(0) \quad \dot{\mathbf{v}}_{s,i}^H(0) \quad \dots \quad \mathbf{v}_{s,i}^{(n)H}(0) \right]$  and  $\mathbf{z}^H(t) =$

$\left[ \mathbf{x}^H(t) \quad t\mathbf{x}^H(t) \quad \dots \quad \frac{t^n}{n!} \mathbf{x}^H(t) \right]$ . Substituting the extended

sample covariance estimate [1] for the expectation, we have:

$$\mathbf{t}_{s,i}^H \hat{\mathbf{R}}_E(t) \mathbf{t}_{s,i} \quad (26)$$

where

$$\hat{\mathbf{R}}_E(t) = \frac{1}{K} \sum_{k=t-\frac{K-1}{2}}^{t+\frac{K-1}{2}} \mathbf{z}(k)\mathbf{z}^H(k) \quad (27)$$

This suggests the following estimation and tracking procedure. First, find the principal components of the extended sample covariance matrix  $\hat{\mathbf{R}}_E(t)$ . Next partition them into

$\mathbf{v}_{s,i}(0)$ ,  $\dot{\mathbf{v}}_{s,i}(0)$ ,  $\dots$  and  $\mathbf{v}_{s,i}^{(n)}(0)$ . Finally use (21) (truncated beyond the  $n^{\text{th}}$  term) to estimate  $\mathbf{v}_{s,i}(t)$ .

#### Example:

Using the signal model from Figure 4, we calculated the sample covariance matrix (i.e., equation (3) with  $K = 21$ ) and its principal eigenvector at each  $t$ . Then, at each  $t$ , we computed the angle between this eigenvector and the principal eigenvector of  $\mathbf{R}(t)$ . This is shown as the solid curve in Figure 5.

Next, we examined the procedure of (26) and (21) (with,  $K = 21$ , and  $n = 20$ ). At each  $t$ , an estimate was formed by numbering the snapshots (surrounding the  $t^{\text{th}}$  snapshot) from -10 ... 10. Then the procedure of (26) and (21) was used to estimate  $\mathbf{v}_{s,i}(0)$ . This, in turn, was used directly as our estimate for the subspace at time  $t$ . As above, the angle between our subspace and the principal eigenvector of the true covariance matrix was computed at each  $t$ . Figure 5 compares the two methods. Observe that the Taylor-based method provided a superior estimate of the signal subspace in this highly nonstationary signal environment.

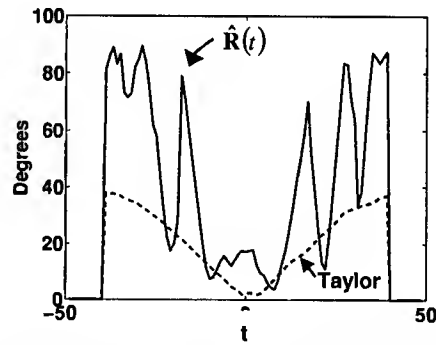


Figure 5. Angle between true and tracked subspaces

## 4. SUMMARY

This paper has proposed a methodology that we have called "Taylor series adaptive processing." In this methodology, quantities of interest are expanded in terms of their Taylor series, and the terms of the series are calculated from the data. This provides a means to efficiently track, compress, and/or interpolate quantities (e.g., adaptive statistics), as well as improved estimates in nonstationary environments.

[1] S. D. Hayward, "Adaptive Beamforming for Rapidly Moving Arrays," 1996 CIE Int. Conf. on Radar, pp. 480-483.

# Adaptive Bayesian Signal Processing – A Sequential Monte Carlo Paradigm \*

Xiaodong Wang<sup>†</sup> and Rong Chen<sup>‡</sup> and Jun S. Liu<sup>§</sup>

## Abstract

We provide a general framework for using Monte Carlo methods in dynamic systems and discuss its wide applications in adaptive signal processing. All of these methods are partial combinations of three ingredients: important sampling and resampling, rejection sampling and Markov chain iterations. Examples from target tracking and digital communication applications are provided to demonstrate the effectiveness of these novel statistical signal processing techniques.

## 1. BACKGROUND ON SEQUENTIAL MONTE CARLO METHODS

In this section, the general framework of sequential Monte Carlo methods for updating a dynamic system is described. Of particular interest is the mixture Kalman filtering technique for on-line estimation of conditional linear dynamic models, which is especially useful for adaptive Bayesian signal processing in time-varying, nonlinear and non-Gaussian environment.

### 1.1. Sequential Monte Carlo Filtering

Consider the following dynamic system modeled in a state-space form as

$$\begin{aligned} \text{state equation} \quad x_t &= f_t(x_{t-1}, u_t) \\ \text{observation equation} \quad y_t &= g_t(x_t, v_t) \end{aligned} \quad (1)$$

where  $x_t$ ,  $y_t$ ,  $u_t$  and  $v_t$  are respectively the state variable, the observation, the state noise and the observation noise at time  $t$ . Denote  $X_t = (x_0, x_1, \dots, x_t)$  and

\*This work was supported in part by the Interdisciplinary Research Initiatives Program, Texas A&M University. R. Chen was supported in part by the U.S. National Science Foundation (NSF) under grant DMS-9626113 and grant DMS-9982846. X. Wang was supported in part by the NSF grant CAREER CCR-9875314. J.S. Liu was supported in part by the NSF grant DMS-9803649.

<sup>†</sup>Department of Electrical Engineering, Texas A&M University, College Station, TX 77843.

<sup>‡</sup>Department of Statistics, Texas A&M University, College Station, TX 77843.

<sup>§</sup>Department of Statistics, Stanford University, Stanford, CA 94305.

$Y_t = (y_0, y_1, \dots, y_t)$ . Suppose that at time  $t$ , we are interested in making some inference about the state variable  $x_t$  based on  $Y_t$ , which is essentially computing  $E\{h(x_t)|Y_t\} = \int h(x_t)p(x_t|Y_t)dx_t$ , for some function  $h(\cdot)$ . In most cases an exact evaluation of this expectation is analytically intractable, due to the complexity of such a dynamic system. Monte Carlo methods provide a viable alternative to facilitate such computations. Specifically, if we can draw  $m$  random samples  $\{x_t^{(j)}\}_{j=1}^m$  from the distribution  $p(x_t|Y_t)$ , then we can approximate  $E\{h(x_t|Y_t)\}$  by

$$E\{h(x_t) | Y_t\} \cong \frac{1}{m} \sum_{j=1}^m h(x_t^{(j)}). \quad (2)$$

Very often directly sampling from  $p(x_t|Y_t)$  is infeasible, but drawing from some *trial* distribution  $q(x_t|Y_t)$  is easy. Suppose that a set of samples  $\{x_t^{(j)}\}_{j=1}^m$  are generated instead from a trial distribution  $q(x_t|Y_t)$ . To utilize these samples for making inferences about the target distribution  $p(x_t|Y_t)$ , the samples have to be weighted by the following weights

$$w_t^{(j)} = \frac{p(x_t^{(j)} | Y_t)}{q(x_t^{(j)} | Y_t)}, \quad j = 1, 2, \dots, m. \quad (3)$$

Then the inference  $E\{x_t|Y_t\}$  can be approximated by

$$E_p\{h(x_t) | Y_t\} \cong \frac{1}{m} \sum_{j=1}^m h(x_t^{(j)})w_t^{(j)}. \quad (4)$$

The pair  $(x_t^{(j)}, w_t^{(j)})$ ,  $j = 1, \dots, m$ , is called a *properly weighted sample* with respect to the distribution  $p(x_t|Y_t)$ .

To implement Monte Carlo for such a system, random samples drawn from  $p(x_t|Y_t)$  are needed at any time  $t$ . A sequential Monte Carlo filter (MCF) for updating the dynamic system (1) involves generating properly weighted samples  $\{(x_t^{(j)}, w_t^{(j)})\}_{j=1}^m$  for  $p(x_t|Y_t)$  at time  $t$ , based on the properly weighted samples  $\{(x_{t-1}^{(j)}, w_{t-1}^{(j)})\}_{j=1}^m$  for  $p(x_{t-1}|Y_{t-1})$  at time  $(t-1)$ ,

according to the following algorithm [Liu and Chen (1998)].

FOR  $j = 1, \dots, m$  DO

1. Draw a sample  $x_t^{(j)}$  from a trial distribution  $q(x_t | X_{t-1}^{(j)}, Y_t)$ ;
2. Compute the importance weight  $w_t^{(j)} = w_{t-1}^{(j)} \cdot p(X_t^{(j)} | Y_t) / [p(X_{t-1}^{(j)} | Y_{t-1}) q(x_t^{(j)} | X_{t-1}^{(j)}, Y_t)]$ .

The samples and weights at time  $t$  are used to approximate the inference  $E\{h(x_t) | Y_t\}$  using (4). It can be shown that the weighted samples generated by this algorithm are unbiased, i.e.,

$$E \left\{ h(x_t^{(j)}) w_t^{(j)} \right\} = E\{h(x_t) | Y_t\}. \quad (5)$$

Hence by the law of large numbers,

$$\frac{1}{m} \sum_{j=1}^m h(x_t^{(j)}) w_t^{(j)} \xrightarrow{a.s.} E\{h(x_t) | Y_t\},$$

as  $m \rightarrow \infty$ . (6)

There are many important issues regarding the design and implementation of a sequential MCF, such as the choice of the trial distribution  $q(\cdot)$ , and the use of *resampling*. Specifically, the most efficient choice of the trial distribution  $q(x_t | X_{t-1}^{(j)}, Y_t)$  for the state space model (1) has the following form

$$\begin{aligned} q(x_t | X_{t-1}^{(j)}, Y_t) &= p(x_t | X_{t-1}^{(j)}, Y_t) \\ &\propto p(y_t | x_t) p(x_t | x_{t-1}^{(j)}), \end{aligned} \quad (7)$$

For this trial distribution, the important weight is updated according to

$$\begin{aligned} w_t^{(j)} &= w_{t-1}^{(j)} \cdot \frac{p(X_t^{(j)} | Y_t)}{p(X_{t-1}^{(j)} | Y_{t-1}) p(x_t^{(j)} | X_{t-1}^{(j)}, Y_t)} \\ &\propto w_{t-1}^{(j)} \cdot p(y_t | x_{t-1}^{(j)}), \end{aligned} \quad (8)$$

### 1.2. The Mixture Kalman Filter

Many dynamic system models belong to the class of conditional dynamic linear models (CDLM) of the form

$$\begin{aligned} x_t &= F_{\lambda_t} x_{t-1} + G_{\lambda_t} u_t, \\ y_t &= H_{\lambda_t} x_t + K_{\lambda_t} v_t, \end{aligned} \quad (9)$$

where  $u_t \sim \mathcal{N}(0, I)$ ,  $v_t \sim \mathcal{N}(0, I)$  (Here  $I$  denotes an identity matrix.), and  $\lambda_t$  is an indicator random variable. The matrices  $F_{\lambda_t}$ ,  $G_{\lambda_t}$ ,  $H_{\lambda_t}$  and  $K_{\lambda_t}$  are known constant matrices given  $\lambda_t$ . It is apparent that in CDLM, for a given trajectory of the indicator  $\lambda_t$ , the system is linear and Gaussian, for which the Kalman filter provides the complete statistical characterization of the system dynamics. Recently a novel sequential Monte Carlo

method, the mixture Kalman filter (MKF), was proposed for on-linear filtering and prediction of CDLM, which exploits the conditional Gaussian property and utilizes a marginalization operation to improve the algorithmic efficiency. The MKF samples in the indicator space and uses a mixture of Gaussian distribution to represent the target distribution. Compared with other sequential MCF methods, substantial performance gains can be achieved by the MKF.

Denote  $Y_t = (y_0, y_1, \dots, y_t)$  and  $\Lambda_t = (\lambda_0, \lambda_1, \dots, \lambda_t)$ . Specifically, the MKF uses the properly weighted discrete samples  $\{(\Lambda_t^{(j)}, w_t^{(j)})\}_{j=1}^m$  to represent  $p(\Lambda_t | y_t)$ , and then it uses a random mixture of Gaussian distributions  $\sum_{j=1}^m w_t^{(j)} \mathcal{N}(\mu_t^{(j)}, \Sigma_t^{(j)})$  to represent the target distribution  $p(x_t | Y_t)$ , where  $\kappa_t^{(j)} \triangleq [\mu_t^{(j)}, \Sigma_t^{(j)}]$  is obtained by implementing a Kalman filter for the given sample trajectory  $\Lambda_t^{(j)}$ . The MKF for updating the CDLM involves generating random samples  $\{(\Lambda_t^{(j)}, \kappa_t^{(j)}, w_t^{(j)})\}_{j=1}^m$  at time  $t$ , based on the samples  $\{(\Lambda_{t-1}^{(j)}, \kappa_{t-1}^{(j)}, w_{t-1}^{(j)})\}_{j=1}^m$  at time  $(t-1)$ , according to the following algorithm [Chen and Liu (2000)]:

FOR  $j = 1, \dots, m$  DO

1. Draw a sample  $\lambda_t^{(j)}$  from a trial distribution  $q(\lambda_t | \Lambda_{t-1}^{(j)}, \kappa_{t-1}^{(j)}, Y_t)$ ;
2. Run a one-step Kalman filter based on  $\lambda_t^{(j)}$ ,  $\kappa_{t-1}^{(j)}$ , and  $y_t$ ;
3. Compute the weight  $w_t^{(j)} = w_{t-1}^{(j)} \cdot p(\Lambda_t^{(j)}, \lambda_t^{(j)} | Y_t) / [p(\Lambda_{t-1}^{(j)} | Y_{t-1}) q(\lambda_t^{(j)} | \Lambda_{t-1}^{(j)}, \kappa_{t-1}^{(j)}, Y_t)]$ .

The MKF can be extended to handle the so-called partial CDLM, where the state variable has a linear component and a nonlinear component.

## 2. EXAMPLES

### 2.1. Target tracking

Designing a sophisticated target tracking algorithm is an important task for both civilian and military surveillance systems, particularly when a radar, sonar, or optical sensor is operated in the presence of clutter or when innovations are non-Gaussian (Bar-Shalom and Fortmann, 1988). We show an example of target tracking with maneuvering.

This situation can be modeled as follows:

$$\begin{aligned} x_t &= H x_{t-1} + F u_t + W w_t \\ y_t &= G x_t + V v_t \end{aligned}$$

where  $u_t$  is the maneuvering acceleration. Here we consider an example of Bar-Shalom and Fortmann (1988) in which a two-dimensional target's position is sampled every  $T = 10$ s. The target moves in a plane with constant course and speed until  $k = 40$  when it starts a slow

90° turn which is completed in 20 sampling periods. A second, fast, 90° turn starts at  $k = 61$  and is completed in 5 sampling times.

$$H = \begin{pmatrix} 1 & 0 & 10 & 0 \\ 0 & 1 & 0 & 10 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}; G = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix};$$

$$F = \begin{pmatrix} 5 & 0 \\ 0 & 5 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}; W = \sigma_w^2 \begin{pmatrix} 5 & 0 \\ 0 & 5 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}; V = \sigma_v^2 \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix};$$

The slow turn is the result of acceleration inputs  $u_t^x = u_t^y = 0.075$  ( $40 < t \leq 60$ ), and the fast turn is from  $u_t^x = -u_t^y = -0.3$  ( $61 < t \leq 65$ ). Other  $u_t$ 's are zero (i.e. no maneuvering).

To apply the MKF to this application, we need to specify prior structure of  $u_t$ . First, we assume that maneuvering can be classified into several categories, indicated by an indicator. In particular, we assume a three level model,  $I_t = 0$  indicates no maneuvering ( $u_t = 0$ ), and  $I_t = 1$  and 2 indicate slow and fast maneuvering, respectively, ( $u_t \sim N(0, \sigma_1^2)$ ,  $\sigma_1^2 < \sigma_2^2$ ). In this study we used  $\sigma_1^2 = 1$  and  $\sigma_2^2 = 36$ . We also specify transition probabilities  $P(I_t = j | I_{t-1} = i) = p_{ij}$  for the maneuvering status. Specifically, we assume  $p_{ii} = 0.8$  and  $p_{ij} = 0.1$  for  $i \neq j$  (i.e. it is more likely to stay in a particular maneuvering state than to change the maneuvering state). Second, there are different ways of modeling the serial correlation of the  $u_t$ . Here we assume a multi-level white noise model, as in Bar-Shalom and Fortmann (1988), where the  $u_t$  are assumed independent, given the indicator. This is the easiest but not a very realistic model. Other possible models are currently under investigation.

In Figure 1 we present the root mean square errors of the MKF estimates of the target position for 50 simulated runs. Comparing our result with that of Bar-Shalom and Fortmann (1988, pp 143) who used the traditional detection-and-switching method, we see a clear advantage of the proposed MKF.

## 2.2. Digital Signal Extraction in Fading Channels

Many mobile communication channels can be modeled as Rayleigh flat-fading channels, which have the following form:

$$\begin{aligned} \text{State Equations: } & \begin{cases} x_t = Fx_{t-1} + Ww_t \\ \alpha_t = Gx_t \\ s_t \sim p(\cdot | s_{t-1}) \end{cases} \\ \text{Observation Equation: } & y_t = \alpha_t s_t + Vv_t \end{aligned}$$

where  $s_t$  are the input digital signals (symbols),  $y_t$  are the received complex signals, and  $\alpha_t$  are the unobserved (changing) fading coefficients. Both  $w_t$  and  $v_t$  are complex Gaussian with identity covariance matrices. This

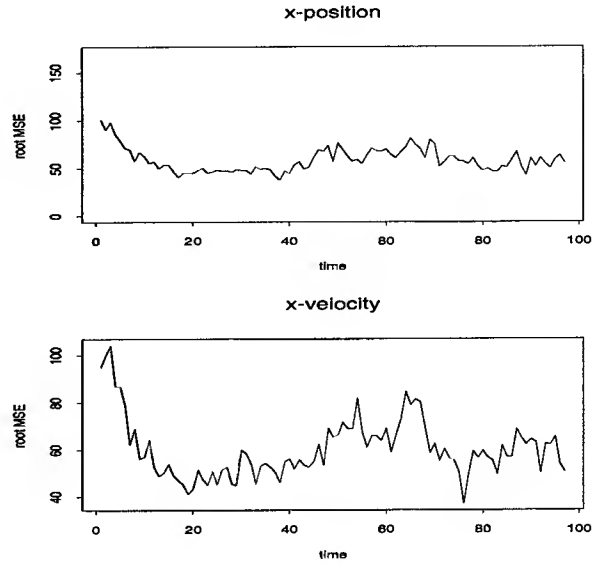


Figure 1. The root MSE's of the  $x$ -position and  $x$ -direction velocity of 50 runs of the MKF for a simulated two-dimensional target moving system with maneuvering.

system is a clearly a PCDLM. Given the input signals  $s_t$ , the system is linear in  $x_t$  and  $y_t$ . Consider binary input signals  $s_t = \{1, -1\}$ . The fading coefficient takes complex values, with independent real and imaginary parts following the same state equation. Both of the real and the imaginary parts of  $\alpha_t$  follow an ARMA(3,3) process

$$\begin{aligned} & \alpha_t - 0.9391\alpha_{t-1} + 2.8763\alpha_{t-2} - 2.9372\alpha_{t-3} \\ & = 0.0376e_t + 0.1127e_{t-1} + 0.1127e_{t-2} + 0.0376e_{t-3} \end{aligned}$$

where  $e_t \sim N(0, 0.01^2)$ . In the communication literature, this is called a (lowpass) Butterworth filter of order 3 with cutoff frequency 0.01. It is normalized to have a stationary variance 1.

We are interested in estimating the differential code  $d_t = s_t s_{t-1}$ . Figure 2 shows the bit error rate of different signal to noise ratios (SNR), using EMKF, the differential detection  $\hat{d}_t = \text{sign}(\text{real}(y_t y_{t-1}^*))$  and a lower bound. The lower bound is obtained using the true fading coefficients  $\alpha_t$  and  $\hat{d}_t = \text{sign}(\text{real}(\alpha_t^* y_t y_{t-1}^* \alpha_{t-1}))$ . The Monte Carlo sample size  $m$  was 100 for the MKF. We also include the result of a delayed estimation, in which  $s_t$  is estimated using the samples  $s_t^{(j)}$  generated by the MKF, and the weight  $w_{t+1}^{(j)}$  at time  $t+1$  (Liu and Chen 1998). This delayed estimation is able to utilize the substantial information contained in the future information  $y_{t+1}$ , and hence is more accurate due to the strong memory in the fading channel.

We can see that the simple differential detection works very well in low SNR cases and no significant improvement can be expected. However, it has an apparent bit error rate floor for high SNR cases. The MKF managed to break that floor, by using the structure of the fading coefficients. The details of this treatment can be found in Chen, Wang and Liu (2000).

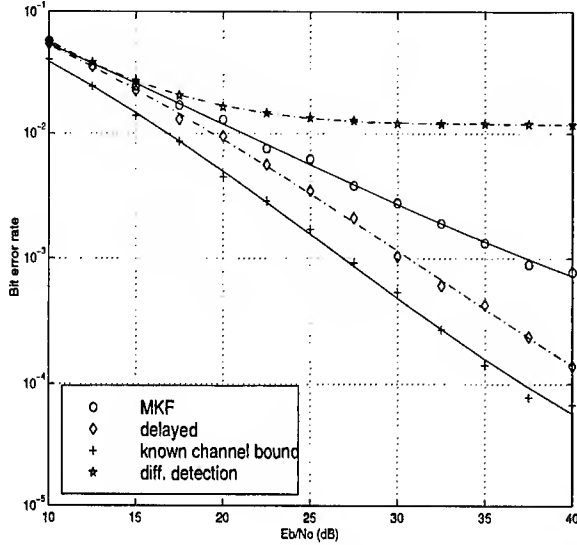


Figure 2. The bit error rate of extracting differential binary signals from a fading channel using differential detection, the MKF and the delayed MKF. A lower bound that assumes the exact knowledge of the fading coefficients is also shown.

### 2.3. Blind Deconvolution.

Consider the following system in digital communication

$$y_t = \sum_{i=1}^q \theta_i s_{t-i} + \varepsilon_t,$$

where  $s_t$  is a discrete process taking values on a known set  $S$ . In a blind deconvolution problem,  $s_t$  is to be estimated from the observed signals  $\{y_1, \dots, y_t\}$ , without knowing the channel coefficients  $\theta_i$ . This system can be formulated as a PCDLM. Let  $\theta_t = (\theta_{t1}, \dots, \theta_{tq})$  and  $x_t = (s_t, \dots, s_{t-q})'$ . We can define

$$\begin{aligned} \text{State Equation: } & \begin{cases} \theta_t = \theta_{t-1} \\ x_t = H x_{t-1} + W s_t \end{cases} \\ \text{Observation equation: } & y_t = \theta_t x_t + \varepsilon_t \end{aligned}$$

where  $H$  is a  $q \times q$  matrix with lower off-diagonal element being one and all other elements being zero and  $W = (1, 0, \dots, 0)'$ . In this case, the unknown system

coefficients are part of the state variable, and are linear conditional on the digital signal  $x_t$ . Liu and Chen (1995) studied this problem with a procedure which is essentially an extended MKF. This PCDLM formulation can be easily extended to deal with a blind deconvolution problem with time-varying system coefficients. In figure 3 we plot the channel estimates as a function of time for a static 4-tap complex ISI channel. It is seen that the channel can be tracked quickly.

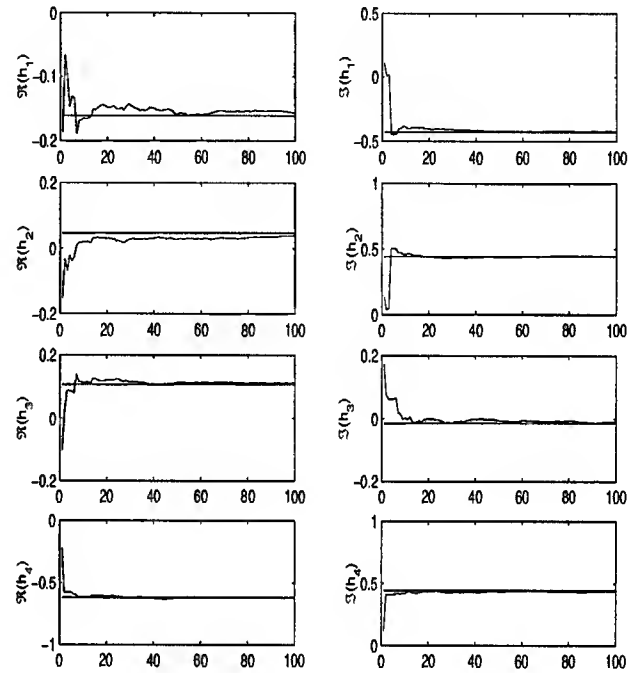


Figure 3. Convergence of the blind equalizer based on sequential imputation.

### REFERENCES

- Bar-Shalom, Y. and Fortmann, T.E. (1988) *Tracking and Data Association*, Academic Press: Boston
- Chen, R and Liu, J.S. (2000) Mixture Kalman Filters. To appear in *Journal of Royal Statistical Society (B)*.
- Chen, R, Wang, X, and Liu, J.S. (2000) Adaptive Joint Detection and Decoding in Flat-Fading Channels via Mixture Kalman Filtering. To appear in *IEEE Transactions on Information Theory*, Sept. 2000.
- Liu, J.S. and Chen, R. (1995) Blind deconvolution via sequential imputations. *J. Amer. Statist. Assoc.*, **90**, 567-576.
- (1998) Sequential Monte Carlo methods for dynamic systems. *J. Amer. Statist. Assoc.*, **93**, 1032-1044

# QQ-PLOT BASED PROBABILITY DENSITY FUNCTION ESTIMATION

Z. Djurovic<sup>\*1</sup>, B. Kovacevic<sup>\*\*</sup>, V. Barroso<sup>\*</sup>

<sup>\*</sup> Instituto Superior Técnico - Instituto de Sistemas e Robótica, Lisbon, Portugal

<sup>\*\*</sup> Faculty of Electrical Engineering, University of Belgrade, Yugoslavia

## ABSTRACT

We present a new algorithm for the estimation of probability density functions (pdfs). This finds a large number of applications in the context of statistical signal processing problems, such as detection, estimation, filtering or pattern recognition and classification. Our approach relies on the QQ-plot technique. The estimates of the first and second order statistics of the observed random data are used together with a suboptimal piecewise linear approximation of the QQ-plot, yielding a new class of pdfs estimators. We describe the algorithm and test it in comparison with other techniques, showing that our approach provides better results.

## 1. INTRODUCTION

In most signal processing problems related to communications and radar/sonar systems, such as detection, estimation, filtering or pattern recognition and classification, the observed signals are corrupted by noise. In those problems, the measurement signals can often be expressed as the sum of the information bearing signal and the measurement noise. Usually, the noise statistics or its distribution are assumed known. In many situations, due to a diversity of well established modeling issues, the measurement noise is assumed Gaussian. Based on this, optimum processing techniques are derived and the properties of the resulting algorithms can be easily studied. However, if the actual probability density function (pdf) of the noise is not Gaussian, or if its statistics are unknown, the performance of those algorithms will degrade significantly, even if the "distance" between the actual noise distribution and the Gaussian one is small. To overcome noise modeling mismatches, and in order to recover the overall performance of the optimum processors, efficient estimators of the noise pdf must be used. In this paper, we will introduce a new approach to the pdf estimation problem, which is based on the QQ-plot technique [3].

## 2. THE QQ-PLOT TECHNIQUE

The probability density function (pdf) of a random process can be estimated from a sequence of samples. Such estimates are required to determine the conditional rate of failure in reliability theory or the decision functions in unsupervised pattern classification problems, and in adaptive filtering problems [1,2,5]. One way of finding the appropriate solution is to look for functional approximations as

$$p(x) \approx \hat{p}(x) = \sum_{i=1}^n c_i \phi_i(x) = c^T \phi(x) \quad (1)$$

where  $c$  is a  $n$ -vector of unknown coefficients,  $(\cdot)^T$  denotes the transpose operator, and  $\phi_i(x)$  is a set of known functions [2]. The problem is to find the vector  $c$  that minimizes a suitable function of the error  $\epsilon(x) = p(x) - c^T \phi(x)$ .

The proposed method is related to the well known QQ-plot statistical technique, which is described here. Let us consider the samples  $\{x_i\}, i=1, \dots, I$  of a random variable  $X$  with pdf  $p(x)$ . By ranking the samples  $\{x_i\}$  we obtain the nondecreasing sequence  $\{y_i\}, i=1, \dots, I$ , where  $y_i \leq y_j$  for any  $(i, j): i \leq j$ . The probability that some observation  $y$  will have rank  $I$  in the ordered sequence  $\{y_i\}$  is given by [2,4]

$$P(i/y) = \binom{I-1}{i-1} P^{i-1}(y) [1 - P(y)]^{I-i} \quad (2)$$

where  $P(\cdot)$  is the distribution function corresponding to the given pdf  $p(\cdot)$ . Then, the conditional expectation  $m_{i/y}$  and the conditional variance  $\sigma_{i/y}^2$  of the random variable  $i$  given the sample  $y$  are, respectively, given by

<sup>1</sup> The work of the first author was partially supported by the Portuguese Foundation for Science and Technology grant PRAXIS\_XXI/BPD/22107/99.

$$\begin{aligned} m_{i/y} &= E\{i/y\} = 1 + (I-1)P(y) \\ \sigma_{i/y}^2 &= E\{i^2/y\} - E^2\{i/y\} = (I-1)P(y)[1-P(y)] \end{aligned} \quad (3)$$

It should be noted that  $\sigma_{i/y}^2$  is small, meaning that the rank  $i$  of the random variable  $y$  is in the vicinity of  $m_{i/y}$ , i.e.,  $i \approx m_{i/y}$ . Then, it follows from (3) that the relation between  $i$  and  $P(y)$  is approximately linear. A plot of the ordered samples  $y_i$  versus the quantity  $P^{-1}(\rho_i)$ , where  $P^{-1}(\cdot)$  is the inverse of  $P(\cdot)$  and  $\rho_i = (i-1)/(I-1)$ , is the QQ-plot [3]. By taking  $m_{i/y} \approx i$  in (3), one obtains the QQ-plot relations

$$y_i = P^{-1}(\rho_i) \approx P^{-1}(r_i); \quad r_i = (i-0.5)/I; i=1,2,\dots,I \quad (4)$$

Thus, if the QQ-plot (4) is fairly linear, then it indicates that the samples have the same distribution  $P(\cdot)$ , even in the tails. Relation (4) can, therefore, be expressed in the linear regression form

$$y_i = m + \sigma P_0^{-1}(r_i) = m + \sigma \bar{r}_i \quad (5)$$

where  $P_0(\cdot)$  is an arbitrary standard distribution, i.e., with zero mean and unit variance, which generates the random variables  $\bar{r}_i = (y_i - m)/\sigma$ , with  $m = E\{y\}$  and  $\sigma^2 = E\{(y-m)^2\}$ . Starting from (5), one can estimate the unknown parameters  $m$  and  $\sigma$  using the least-squares algorithm [4],

$$\begin{bmatrix} \hat{m} \\ \hat{\sigma} \end{bmatrix} = (\Sigma^T \Sigma)^{-1} \Sigma^T Y, \quad (6)$$

where  $\Sigma^T = \begin{bmatrix} 1 & 1 & \dots & 1 \\ \bar{r}_1 & \bar{r}_2 & \dots & \bar{r}_I \end{bmatrix}$  and  $Y^T = [y_1 y_2 \dots y_I]$ .

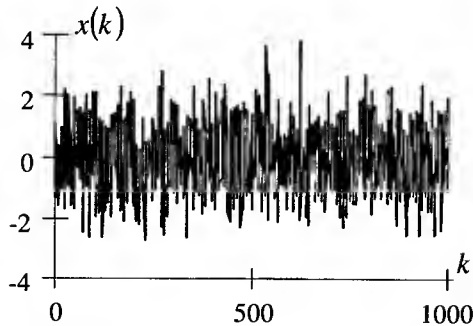


Fig.1. Standard Gaussian random variable samples

In order to illustrate the technique, we generated 1000 samples of the standard Gaussian random variable

( $m=0, \sigma=1$ ), see Fig.1. The corresponding QQ-plot is presented in Fig. 2. By applying relation (6), we obtained statistics estimates  $\hat{m}=0.0057$  and  $\hat{\sigma}=1.0154$ . The properties of these estimators can be found in [4].

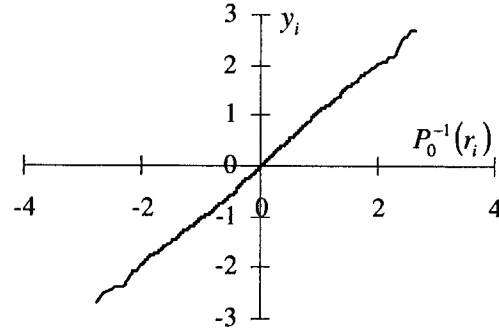


Fig. 2. Corresponding QQ-plot

### 3. PROPOSED ALGORITHM

In this section, the idea of obtaining the pdf's estimate  $\hat{p}$  in (1) using the QQ-plot method is elaborated. To do this, we have to segregate  $n$  linear parts of the given QQ-plot, resulting in the partition of the data set in  $n$  mutually exclusive subsets with cardinal numbers  $c_i$ ,  $i=1,\dots,n$ . Then, the term  $\phi_i(\cdot)$  in (1) can be interpreted as an approximation of the conditional pdf, given the data points from the  $i$ -th subset, while the coefficients  $c_i$  represent the a priori probability of the observations to take values in the  $i$ -th subset. Furthermore, such obtained linear segments are defined by (5), and are uniquely determined by the first and second order moments  $m_i$  and  $\sigma_i$  of  $\phi_i(\cdot)$ , whose respective estimates  $\hat{m}_i$  and  $\hat{\sigma}_i$  are given in (6). The key step in the proposed procedure is to generate an optimal piece-wise linear approximation of the QQ-plot. Of course, this is a nonclassical optimization problem, which is tractable only by numerical methods. Therefore, we propose a heuristic procedure that yields a suboptimal solution.

Let  $A$  denote the set of all piece-wise linear functions  $a(\cdot)$  with a number of linear segments  $n_a$ , whose domain is the interval of real numbers  $[P_0^{-1}(r_1), P_0^{-1}(r_I)]$ . The goal is to find a function  $a^* \in A$  that satisfies:

$$a^* = \arg \min_{a \in A} J(a) = \arg \min_{a \in A} \{ \text{dist}(a, qq) + \gamma n_a \}; \quad (7)$$



where  $dist(a, qq) = \frac{1}{I} \sum_{i=1}^I (a(P_0^{-1}(r_i)) - y_i)^2$  and

$y_i = qq(P_0^{-1}(r_i))$ ,  $i=1,2,\dots,I$ , is the set of nondecreasing measurements,  $\gamma$  being a properly defined parameter. The influence of this parameter to the final solution is great. Namely, for  $\gamma=0$  the final solution  $a^*$  would be a piece-wise linear function with  $(I-1)$  linear segments that would approximate QQ-plot ideally. On the other hand, each of the segments would contain only two terminal measurements, and this represents a small amount of data for proper statistic estimation. On the contrary, a too large  $\gamma$  will cause a solution  $a^*$  with a small number of segments, resulting in a rather bad approximation of the QQ-plot. Thus, parameter  $\gamma$  has to be chosen as a compromise between those two opposite requirements. A suitable choice is  $\gamma \in (0.01, 0.05)$ . Finally, we achieved convergence towards the optimal solution  $a^*(\cdot)$  through the functional sequence  $\{a_j(\cdot)\}$  obtained on the following manner:

**step i)** in the initial step the first member  $a_1(\cdot)$  consists of only one linear segment ( $n=1$ ), determined by the first  $(P_0^{-1}(r_1), y_1)$  and the last point  $(P_0^{-1}(r_I), y_I)$  of the QQ plot;

**step ii)** the member  $a_{j+1}(\cdot)$ ,  $j=1,2,\dots$ , is formed from  $a_j(\cdot)$  using the rule: to each of the linear segments of  $a_j(\cdot)$ , say  $a_{j1}(\cdot), a_{j2}(\cdot), \dots, a_{jk_j}(\cdot)$  compute

$$d(a_{ji}, qq) = \frac{1}{m_{ji}} \sum_{p=1}^{m_{ji}} (a_{ji}(P_0^{-1}(r_p)) - qq(P_0^{-1}(r_p)))^2 = l_i,$$

$i=1,\dots,k_j$ , where  $m_{ji}$  denotes the number of the QQ-plot points covered by the  $i$ -th linear segment in the  $j$ -th iteration. Denote the maximal value with  $l^* = \max_{i=1,\dots,k_j} l_i$ .

Now, divide the segment with corresponding  $l^*$ -measure into two linear subsegments so that these subsegments contain the same number of data points. So obtained piece-wise linear approximation, that contains  $k_j + 1 = k_{j+1}$  segments, forms the member  $a_{j+1}(\cdot)$ ;

**step iii)** if  $dist(a_j, qq) + \gamma n_{a_j} > dist(a_{j+1}, qq) + \gamma n_{a_{j+1}}$ , go to step ii), else finish the algorithm and stop the procedure with  $a^* = a_j$ .

Having finished the construction of the piece-wise linear approximation  $a^*$  of the QQ-plot with  $n$  linear subsegments, the coefficients  $c_i$ ;  $i=1,\dots,n$ , can be

determined as the ratio between the number of data points belonging to the  $i$ -th subset and the total number of data points, while the parameters  $\hat{m}_i$  and  $\hat{\sigma}_i$  are estimated according to (6). So, the pdf approximation  $\hat{p}_i$  can be found in a kernel form

$$\hat{p}(x) = \sum_{i=1}^n c_i \frac{1}{\hat{\sigma}_i} p_0\left(\frac{x - \hat{m}_i}{\hat{\sigma}_i}\right). \quad (8)$$

The choice of  $p_0(\cdot)$  depends on the nature of the data. In the case of symmetric and unimodal pdf of the samples sequence, a logical choice for  $p_0(\cdot)$  should be the normal, uniform or Laplace pdfs. It should be noted that for the normal  $p_0(\cdot)$ , the proposed algorithm represents a Gaussian mixture type estimator [5], while for the uniform  $p_0(\cdot)$ , it can be viewed as a kind of histogram method with variable cells [1].

#### 4. AN EXAMPLE

In order to illustrate the performance of the proposed estimator, we select as an example the Gaussian mixture pdf

$$p(x) = 0.3N(x/-5, 1) + 0.3N(x/-1.5, 1.5) + 0.4N(x/1.5, 1) \quad (9)$$

where  $N(\cdot/a, b)$  denotes the normal pdf with mean  $a$  and variance  $b$ . The performance of the algorithm is shown in Figs. 3, 4.

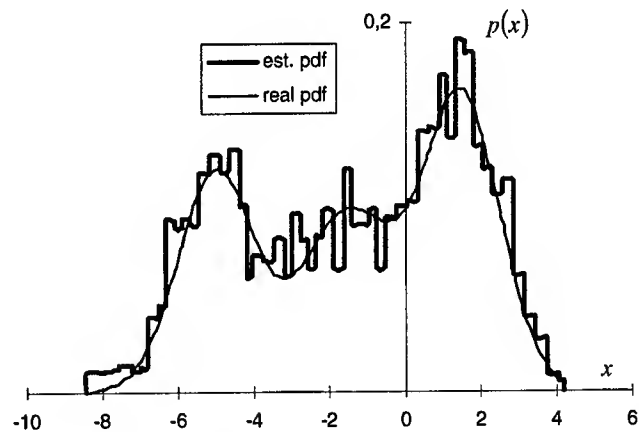


Fig. 3.a: Histogram method with variable cells

To compare our parametric-nonparametric approach to both a nonparametric and a parametric-nonparametric distinct solutions, we also have implemented the histogram method with variable cells [1] and the adaptive

Gaussian mixture method [5]. The results were obtained using  $I=5000$  samples from the pdf (9). Estimation results (Fig. 5) are compared in terms of the integral estimation error (CEE):

$$CEE(I) = \int_{-\infty}^{\infty} (p(x) - \hat{p}_I(x))^2 dx \quad (10)$$

where  $I$  denotes the number of data samples used for estimation. The numerical complexity of the estimators is analyzed and expressed through the number of floating point operations. The corresponding results are plotted in Fig. 6.

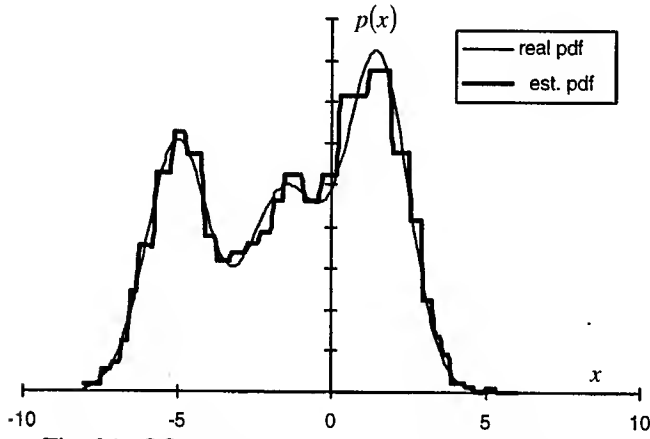


Fig. 3.b: QQ-based estimation with uniform pdf  $p_0$

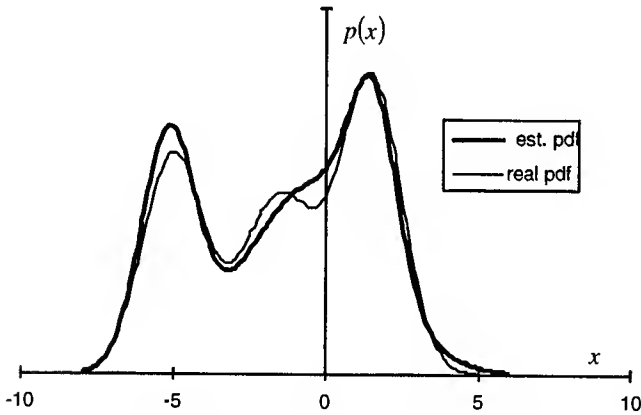


Fig. 4.a: Adaptive Gaussian mixture method

The achieved results show clearly that our algorithm outperforms either the histogram with variable cells or the Gaussian mixture methods. In fact, the estimated pdf's is a rather good approximation of the actual pdf, being better, or at least comparable, than those obtained with the other two methods. Moreover, the choice of the normal pdf  $p_0(\cdot)$  in (8) represents a good compromise between accuracy and numerical complexity of the proposed estimation procedure.

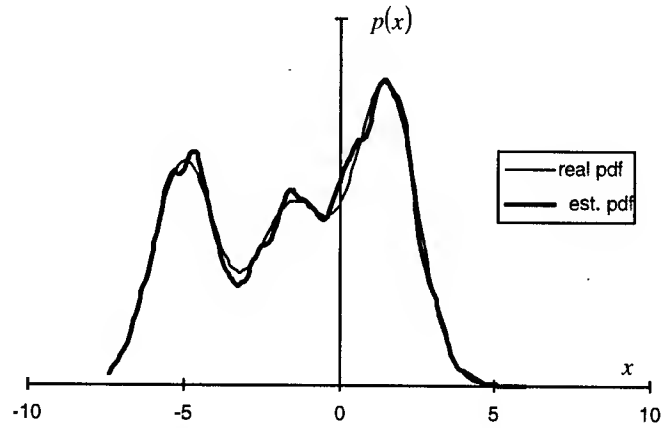


Fig. 4.b: QQ-plot based estimation with normal pdf  $p_0$

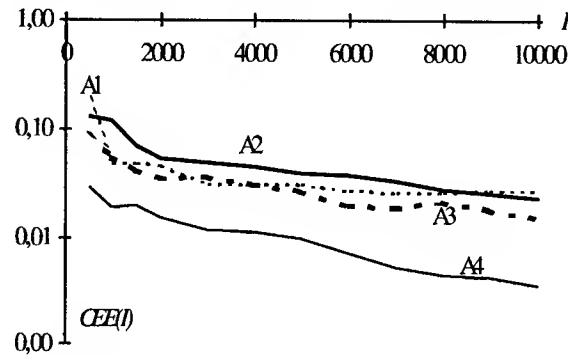


Fig. 5: Comparison of different algorithms in terms of integral estimation error (A1: Histogram with variable cells, A2: Adaptive Gaussian mixture, A3: QQ-plot based estimator with uniform  $p_0$ , A4: QQ-plot based estimator with normal  $p_0$ )

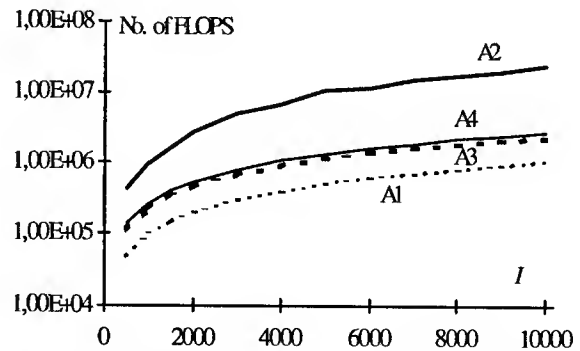


Fig. 6: Comparison of different algorithms in terms of number of floating point operations (A1: Histogram with variable cells, A2: Adaptive Gaussian mixture, A3: QQ-plot based estimator with uniform  $p_0$ , A4: QQ-plot based estimator with normal  $p_0$ )

## 5. CONCLUSION

Inspired by the QQ-plot technique, we developed an algorithm to estimate a pdf from a sequence of samples. The highlights of the presented algorithm are the following: 1) it is described as mixture parametric-nonparametric approach; its nonparametric character implies that we can make a simple assumption about the data, i.e., concerning the function  $p_0(\cdot)$  under which the QQ-plot is constructed; 2) the final pdf estimate result is in the form of an analytical function rather than a numerical function; 3) it is computationally simpler and performs better, or comparably, with respect to similar parametric-nonparametric methods, like the adaptive Gaussian mixture algorithm; 4) it performs better than recursive parametric methods, like the histogram method with variable cells, at the expense of a modest additional computational effort; and 5) in contrast to classical methods, which aim at getting the actual pdf, it only approximates this pdf, just as parametric methods do. The drawback of the proposed approach is its nonrecursivity, which means that all samples have to be stored during the computation of the pdf estimate. Nevertheless, it presents a good compromise between estimation accuracy and numerical complexity, and offers a good alternative to the other methods known from the literature. Also, there are a diversity of possible applications. Whenever it is necessary to estimate probability density functions of the

signals, patterns of clutter, as it is the case in digital communications, pattern recognition or radar/sonar systems, our method can be used. Also, the estimation algorithm can be implemented relying on sliding window techniques, yielding therefore the possibility of operation on slowly time varying scenarios. In these situations, the length of the sliding window has to be a compromise between the desired estimation accuracy and the changing rate of the statistics/distributions of interest.

## 6. REFERENCES

- [1] Fukunaga K., *Introduction to Statistical Pattern Recognition*, Academic Press, New York, 1990.
- [2] Pugatchev V., *Probability Theory and Mathematical Statistics for Engineers*, Pergamon Press, New York, 1984.
- [3] Gnanadesikan R., *Methods for Statistical Data Analysis of Multivariate Observations*, John Wiley, New York, 1977.
- [4] Djurovic Z., Kovacevic B., "QQ-plot Approach to Robust Kalman Filtering", *Int. J. Control*, 61(4):837-857, 1995.
- [5] Priebe C., Marchette D., "Adaptive Mixtures: Recursive Nonparametric Pattern Recognition", *Int. J. Pattern Recognit.*, 24:1197-1209, 1991.

# NONLINEAR SYSTEM INVERSION APPLIED TO RANDOM VARIABLE GENERATION

*A. Pagès-Zamora, M. A. Lagunas, X. Mestre*

Dpt. Teoria del Senyal i Comunicacions, Universitat Politècnica de Catalunya (UPC),  
C/ Gran Capità s/n, Campus Nord UPC, D-5, 08034 Barcelona, Spain  
[alba@gps.tsc.upc.es](mailto:alba@gps.tsc.upc.es)

## ABSTRACT

In this paper, a method to design random variables (rv) generators with the same probability density function (pdf) as a given rv record is presented. The resulting rv generator is a nonlinear system that, when driven by a uniformly distributed rv, provides an output rv with the desired pdf distribution. The analytical description of the desired pdf is not needed; in fact, only a data record of the desired rv is used. Inversion of nonlinear systems and nonlinear system adaptive design are used in this work.

## 1. INTRODUCTION

For simulation purposes, generators of random variables (rv) with a given probability distribution function (pdf) are often needed. For instance, in the model for a time-space radio channel of [1], it is shown that the wave azimuth distribution almost matches a Gaussian pdf, whereas their delay distribution approximately fits an exponential pdf. Nevertheless, the measured pdf could not always properly fit an analytic pdf distribution with an acceptable confidence level and over the entire range. For instance, in this radio channel model, the tails of the measured azimuth distribution are not well fit by a normal pdf.

In this paper, a method to design rv generators with the same pdf distribution as a given rv record is presented. As shown below, the resulting rv generator is a nonlinear system (NLS) that, when driven by a uniformly distributed rv, provides an output rv with the desired pdf distribution. It is important to point out that the analytical description of the desired pdf is not needed; in fact, only a data record of the desired rv is used. As will be shown, NLS inversion and NLS adaptive design are involved in the design.

The paper is organized as follows. In section II, the "whitening" of a rv is presented. That is, we describe a procedure to obtain a uniformly distributed rv from a given rv record with another pdf. In section III, the rv generator design problem leads to an NLS inversion problem, which is solved adaptatively. Section IV presents simulations and, the paper ends with conclusions in section V.

## 2. PDF WHITENING

In [2], a parallelism between the role that a pdf function plays in nonlinear signal processing and the role that the power spectrum density function plays in linear signal processing is presented.

This relationship allows us to solve the problem of pdf whitening (that is, to obtain a uniformly distributed rv from another rv) similarly to whitening the power spectral density of a stochastic process.

The pdf whitening problem involves the design of an NLS system, denoted by  $g[\cdot]$ , that provides a uniformly distributed rv output, denoted hereafter by  $u(n)$  with  $n$  discrete time index, whenever it is driven by a data record  $x(n)$  of a given distribution. Thus, we have,

$$u(n) = g[x(n)]. \quad (1)$$

It is well-known [3] that such a system is,

$$u(n) = g[x(n)] = 2U_0 \left[ F_X(x(n)) - \frac{1}{2} \right] \quad (2)$$

with  $F_X(x)$  the input distribution function and  $U_0$  the output range, i.e.  $u \in [-U_0, U_0]$ . As (2) is monotonically increasing, the relation between the input pdf,  $p_X(x)$ , and the output one,  $p_U(u)$ , is

$$p_U(u) = p_X(x) \left/ \left( \frac{dg(x)}{dx} \right) \right. \quad (3)$$

and it can be stated in the following integral form:

$$\int_{-\infty}^u p_U(\alpha) \cdot d\alpha = \int_{-\infty}^x p_X(\lambda) \cdot d\lambda. \quad (4)$$

Assuming that the input range is finite<sup>1</sup>, i.e.  $x \in [-X_0, X_0]$ , and stating the input pdf function  $p_X(x)$  in terms of the Fourier series approach, expression (4) leads to

$$\frac{u + U_0}{2U_0} = \frac{1}{2X_0} \sum_{k=-\infty}^{+\infty} \psi_X \left( jk \frac{\pi}{X_0} \right) \cdot \int_{-X_0}^x e^{-jk \frac{\pi}{X_0} \lambda} d\lambda \quad (5)$$

with  $\psi_X[jv]$  the characteristic function of the rv  $x$ . Due to the Fourier series periodicity, (5) is valid only for  $u$  and  $x$  values within their respective ranges. It is straightforward to see that (5) leads to the following pdf whitening system

<sup>1</sup> If it were not finite, a truncation of the input range would be assumed with a certain overflow probability

$$u = g[x] = \frac{U_0}{X_0} \left[ x + \sum_{\substack{k=-\infty \\ k \neq 0}}^{+\infty} \psi_X \left( jk \frac{\pi}{X_0} \right) \frac{e^{-jk \frac{\pi}{X_0} x} - (-1)^k}{-jk \frac{\pi}{X_0}} \right]. \quad (6)$$

For practical purposes, the infinite summation in (6) is truncated to  $|k| \leq K$  and the characteristic function can be estimated by the sample estimator. Assuming  $N$  samples of  $x$ , the characteristic function estimate could be,

$$\hat{\psi}_X \left( jk \frac{\pi}{X_0} \right) = \frac{1}{N} \sum_{n=1}^N e^{jk \frac{\pi}{X_0} x(n)} \quad (7)$$

leading to the approximate pdf whitening system,  $\hat{g}[\cdot]$ .

$$\hat{g}[x(n)] = \frac{U_0}{X_0} \left[ x + \sum_{\substack{k=-K \\ k \neq 0}}^{+K} \hat{\psi}_X \left( jk \frac{\pi}{X_0} \right) \frac{e^{-jk \frac{\pi}{X_0} x(n)} - (-1)^k}{-jk \frac{\pi}{X_0}} \right] \quad (8)$$

Important to remark is that, unlike a simpler whitening system consisting, for instance, of the direct estimation of (2), this pdf whitening system allows a recursive computation of  $\hat{\psi}_X(jk\pi/X_0)$  and enables the system to whitening non-stationary rv. Additionally, although outside the scope of this paper, it is worth to point out that the previous pdf whitening system can be generalized to an arbitrary number of rv's (see [2] for details).

In order to show the performance of the presented pdf whitening system, 2000 samples of a normally distributed rv  $x: N(0,1)$  are considered. Let us assume that  $|x| \leq X_0=3$ , i.e. an overflow probability of  $10^{-3}$  is allowed. From the estimated values of the characteristic function  $\hat{\psi}_X(jk\pi/X_0)$  for  $|k| \leq K=10$  (7) and considering  $U_0=5$ , the pdf whitening system (8) is obtained. Figure 1 shows the normalized histogram of an  $x$  data record of 8000 samples and the resulting "whitened"  $u$  samples obtained from (8). As seen, the output rv histogram has a flat shape.

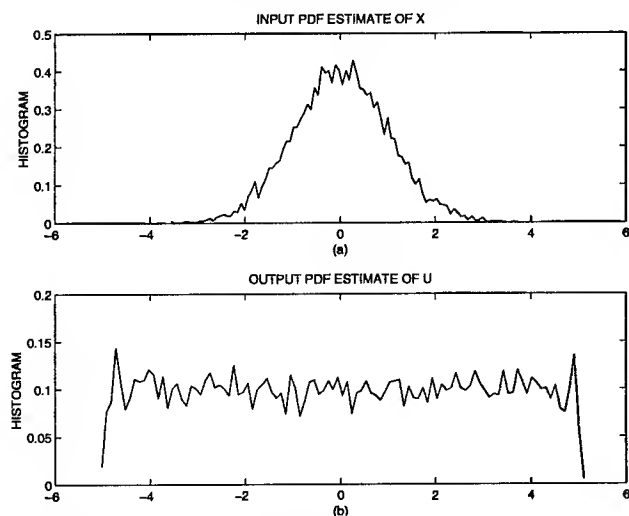


Figure 1. Histograms (8000 samples) of  $x$  (a) and  $u$  (b).

### 3. RV GENERATOR

From here on, we focus on the problem of designing a NLS system whose output has a given pdf function when it is driven by a uniformly distributed rv. Hereafter, such a system will be referred to as an rv generator.

In the previous section, we showed that a data record of a rv  $x$  with a given pdf provides us with a NLS system able to generate a uniformly distributed rv when that NLS system is driven by  $x$ . Consequently, as shown in Figure 2, the design of a  $x$  rv generator system becomes a nonlinear inverse system design problem of the pdf whitening NLS in (8).

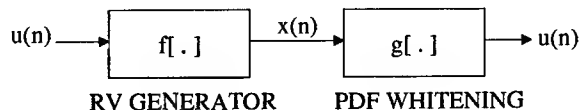


Figure 2. NLS inversion to design the rv generator.

According to (2), the ideal rv generator function  $f[u]$  is

$$f[u] = g^{-1}[u] = F_X \left( \frac{u + U_0}{2U_0} \right). \quad (9)$$

For the sake of comparison, two different NLS designs are considered to model the rv generator: a Volterra model (10) denoted by  $f_V(u)$  and a trigonometric or Fourier model (11) denoted by  $f_F(u)$  [2].

$$\hat{f}_V(u(n)) = \sum_{q=0}^Q a_V(q) \cdot u^q(n) \equiv a_V' \cdot z_V(n) \quad (10)$$

$$\hat{f}_F(u(n)) = a(0) + \sum_{q=1}^Q [a_F(2q) \cdot \cos(2q\omega_0 u(n)) + a_F(2q+1) \cdot \sin((2q-1)\omega_0 u(n))] \equiv a_F' \cdot z_F(n) \quad (11)$$

The linearity of both models with respect to the coefficients enables a vector notation as seen in (10) and (11). The vectors of the nonlinear models are  $a_V$ , the Volterra coefficient vector,  $z_V(n)$  the Volterra functional vector consisting of the powers of  $u(n)$ ,  $a_F$  the Fourier coefficient vector and  $z_F(n)$  the Fourier functional vector consisting of the sine or cosine functions of  $u(n)$ . Also in (11) the so-called principal frequency is defined as  $\omega_0 = \pi/(2X_0)$ . (See [2] for detail about the Fourier model).

As shown in Figure 3, the design of the rv generator can be accomplished in an adaptive manner by means of the so-called Predistortion-LMS (PLMS) [4].

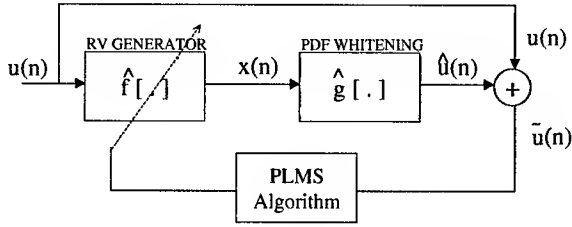


Figure 3. Adaptive design of the rv generator.

The PLMS update of the Volterra or Fourier coefficients follows

$$a(n+1) = a(n) + \frac{\mu}{p(n)} \cdot \tilde{u}(n) \cdot \nabla_x \hat{g}(x) \big|_{x(n)} \cdot z(n), \quad (12)$$

substituting  $a(n)$  for the respective coefficient vector and  $z(n)$  for the respective functional vector, as defined in (10) for the Volterra model and in (11) for the Fourier model. In (12)  $\mu$  is the step-size parameter,  $\tilde{u}(n)$  is the error signal,

$$\tilde{u}(n) = u(n) - \hat{u}(n) \quad (13)$$

and  $p(n)$  is the estimate of the power of the functionals,

$$p(n+1) = \beta \cdot p(n) + (1-\beta) \cdot z'(n) \cdot z(n). \quad (14)$$

The PLMS adaptive algorithm is a gradient algorithm useful in NLS inversion problems because it includes, due to the chain rule, the gradient  $\nabla_x \hat{g}(x)$  of the function to be inverted. From (2), the gradient depends on  $p_X(x)$ .

$$\nabla_x g[x] = 2U_0 \cdot \nabla_x g[F_X(x)] = 2U_0 \cdot p_X(x) \quad (15)$$

For practical purposes, the gradient of (8) can be used directly. Different pdf estimates could be also taken into account to estimate de gradient [2].

The design of the rv generator could have also been performed in reverse order, that is,  $\hat{g}(\cdot)$  could have been put in front of the  $\hat{f}(\cdot)$  in Figure 3. In that case, a least square solution of the NLS model of the rv generator would be feasible because the signal error is linear with the coefficients. The limitation is that, in the reverse order, a large record of  $x$  would be needed, and the objective of the paper is precisely to design a generator of the  $x$  rv from a small record of  $x$ .

#### 4. SIMULATIONS

Two sets of simulations are included. The first one uses the actual characteristic function, whereas in the second simulation only a record of a  $x$  rv is assumed.

First, let us consider a Laplacian rv whose pdf is,

$$p_X(x) = \alpha/2 \cdot e^{-\alpha|x|} \quad (16)$$

with parameter  $\alpha$  set to 1. The pdf whitening system is built from expression (8) with  $U_0=1$  and using the actual samples of the characteristic function for  $K=10$ .

$$\psi_X(jk \frac{\pi}{X_0}) = \frac{\alpha^2}{\alpha^2 + k^2 \frac{\pi^2}{X_0^2}} \quad |k| < K = 10 \quad (17)$$

Due to symmetry of the distribution function, the pdf whitening system and the inverse system both have odd input/output relations. Thus, the Volterra system (10) that models the rv generator only keeps the odd powers, whereas the Fourier model keeps only the sine functionals. Both models consist of 15 coefficients.

A 5000-length data record of a uniformly distributed rv is used in the adaptive design of the rv generator (3). The PLMS parameters (12) (14) are set to  $\mu = 2$  and  $\beta = 0.99$  for both models. The gradient function (15) is computed using the Laplacian pdf function (16). The final relations of  $\hat{f}_V(u)$  and  $\hat{f}_F(u)$  together with the ideal ones (dashed line) are shown in figure (4.a) and (4.b).

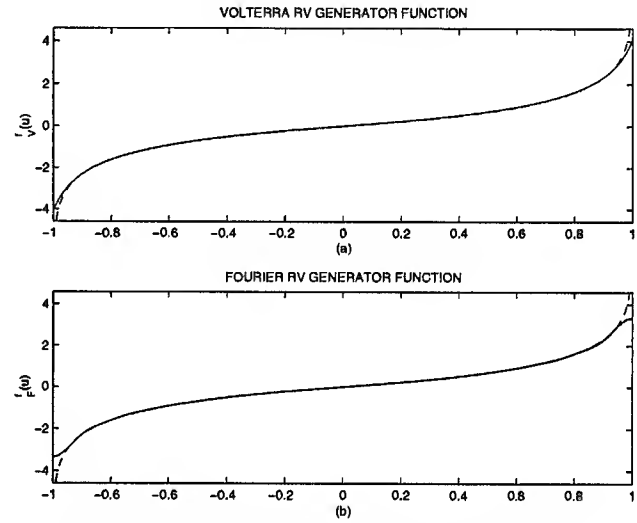
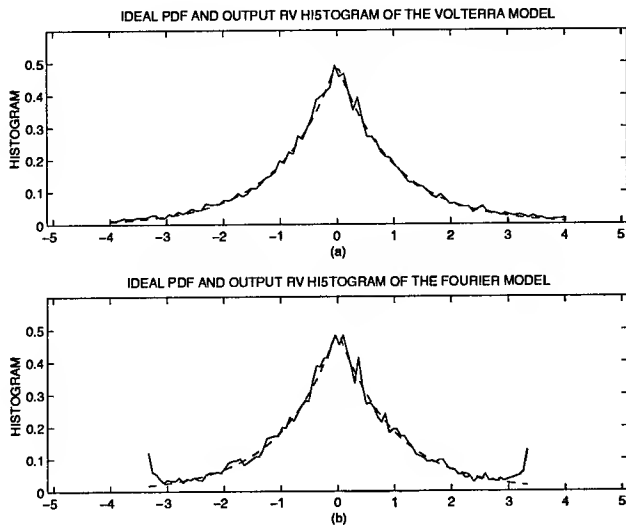


Figure 4. Ideal rv generator in dashed line. In solid line, Volterra (a) and Fourier (b) rv generator functions.

Figure (5) compares the Laplacian pdf (dashed line) to the output rv histogram of the Volterra rv generator (Fig. 5.a) and Fourier rv generator (Fig. 5.b). Both histograms have been computed from  $2 \cdot 10^4$  length data records.

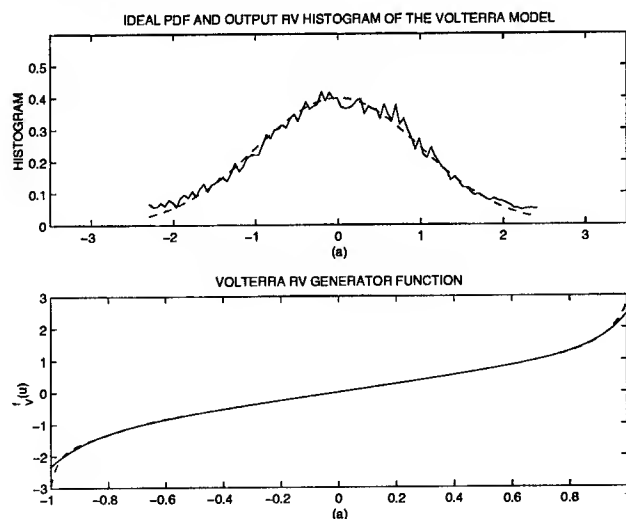
Although not shown, the convergence of the Fourier coefficients is faster than that of the Volterra coefficients, but the Fourier model does not properly fit the tails of the Laplacian pdf function (Fig. 5.b). This is due to the fact that the ideal function  $f[u]$  has a sharp behavior at the boundaries of the input range (see Fig. 4) that the Fourier model does not match properly. In this case, the Volterra model provides better performance for such a NLS design.



**Figure 5.** Laplacian pdf function (dashed line). In the solid line, the output histogram of the Volterra (a) and Fourier (b) rv generator systems.

The second set of simulations uses only 2000 samples of a normal distributed rv  $x: N(0,1)$ . As shown in section 2, this data record allows the design of the pdf whitening system (in this case  $U_0=1$ ). Once the pdf whitening system is obtained, the rv generator can be adaptatively designed using the scheme of Figure 3.

For that purpose, a Volterra system (10),  $\hat{f}_V(u)$ , with  $Q=15$  is considered to model the rv generator. The coefficients are updated with  $2 \cdot 10^4$  samples of a uniformly distributed rv  $u$  and by means of the PLMS adaptive algorithm with  $\mu=2$  and  $\beta=0.99$ . The Fourier series approximation of  $p_X(x)$  ( $K=10$ ) is used to compute gradient function,  $\nabla_x \hat{g}(x)$ .



**Figure 6.** (a) Ideal pdf (dashed line) and histogram of the rv generator output (solid line). (b) Ideal rv generator function (dashed line) and actual rv generator function (solid line).

Figure (6.b) shows the ideal input/output relation of the rv generator system in the dashed line along with the final one achieved by the Volterra system after the adaptive design. Additionally, figure (6.a) shows the actual pdf (dashed line) and the histogram of the Volterra rv generator output using  $2 \cdot 10^4$  samples.

## 5. REMARKS

This paper shows how a nonlinear system that generates a rv with a given pdf can be designed from knowledge only of a data record of such a rv. It has been shown that data records of 2000 samples are large enough to obtain a reliable rv generator system. As a preliminary step, we also presented the design of nonlinear systems that are able to provide a uniformly distributed rv at the output when driven by an input signal with a given pdf.

## 6. ACKNOWLEDGEMENTS

This work has been supported by CICYT: TIC96-0500-C10-01, TIC98-0412, TIC98-0703 and CIRIT: 1998SGR-00081

## 7. REFERENCES

- [1] Pedersen K.I., Mogensen P.E., Fleury B.H., "A Stochastic Model of the Temporal and Azimuthal Dispersion seen at the Base Station in Outdoor Propagation Environments", Submitted to Trans VTC, 1999.
- [2] Pagès-Zamora A, Lagunas M.A, "Fourier Models for Non-Linear Signal Processing", EURASIP Signal Processing, Vol. 76, pp.1-16, June 1999.
- [3] A. Papoulis, Probability, Random variable sand Stochastic Processes, McGraw-Hill, USA, 1965.
- [4] Stonick J.T. *et al.*, "Memoryless Polynomial Adaptive Predistortion", Proc. IEEE-ICASSP95, pp. 981-984, Detroit, Michigan, USA, May 1995.

# THE NUMERICAL SPREAD AS A MEASURE OF NON-STATIONARITY: BOUNDARY EFFECTS IN THE NUMERICAL EXPECTED AMBIGUITY FUNCTION

Robert A. Hedges and Bruce W. Suter

Air Force Research Laboratory IFGC  
525 Brooks Rd.  
Rome, NY 13441-4505  
hedgesr@rl.af.mil, suterb@rl.af.mil

## ABSTRACT

Establishing measures for local stationarity is an open problem in the field of time-frequency analysis. One promising theoretical measure, known as the spread, provides a means for quantifying potential correlation between signal elements. In previous papers we investigated the issue of implementing such a measure for discrete signals. The *numerical spread* was introduced [1] as a means of applying and investigating the techniques previously only studied theoretically.

When implementing such a scheme it became necessary to augment the covariance matrix so that the resulting ambiguity space has a uniform resolution. In this paper we compare three extension schemes: zero padding, circular extension, and edge replication, to determine which provides the best estimate of the numerical spread. Based on our results, we determined that the method of normalized edge replication is least likely to inflate the estimate of the spread.

## 1. INTRODUCTION

One assumption of most signal processing techniques is that the signal is stationary *i.e.*, the statistics do not change over time. Many real data sets are not stationary but can, however, be described as locally stationary, that is they appear stationary over finite time intervals. Since we will not assume access to all orders of statistics we will constrain our discussion to the second order statistics of a signal.

Establishing measures for local stationarity is an open problem in the field of time-frequency analysis. Some desirable properties for such a measure [2]:

- *Quantitative*    • *Measurable*
- *Robust*        • *Analytically Powerful*

One promising theoretical measure, known as the spread, was introduced by W. Kozek [3], and provides

a means for quantifying potential correlation between signal elements.

When implementing such a scheme numerically, it will be shown that the natural way in which the data is defined leads to calculation of FFTs of various lengths. This is undesirable since the resulting spectra will be calculated at different resolutions. This work investigates methods that address the problem of augmenting the covariance matrix such that calculations are made on a grid of constant frequency without introducing other artifacts.

### 1.1. Local Stationarity

Before reviewing the theory of spread, it will be useful to explicitly define local stationarity. A signal is locally stationary if the following conditions hold:

1. The autocovariance has limited variation over some time interval  $T$ .
2. For lags,  $\tau$ , that extend beyond the interval  $T$  the autocovariance needs to be small, preferably zero, to prevent time dependence outside the interval.

$$R_x(t, t + \tau) \approx 0, \text{ for } t \in T \text{ and } t + \tau \notin T.$$

### 1.2. Theoretical Spread

To establish a context for numerical spread we first summarize the theoretical framework for spread of a random process  $x(t)$  defined for continuous  $t$ . To quantify the degree of local stationarity we introduce the (generalized) ambiguity function and the expected (generalized) ambiguity function.

The (generalized) Ambiguity Function (AF) [4] of a deterministic signal  $x(t)$  is given by

$$A_x^{(\alpha)}(\tau, \nu) = \int x(t + (\tfrac{1}{2} - \alpha)\tau) x^*(t - (\tfrac{1}{2} + \alpha)\tau) e^{-i2\pi\nu t} dt.$$



Given a nonstationary random process  $x(t)$  the *expected (generalized) ambiguity function* EAF  $EA_x^{(\alpha)}(t, \nu)$  is the expectation of the AF

$$\begin{aligned} EA_x^{(\alpha)}(\tau, \nu) &= E \left[ \int x(t + (\tfrac{1}{2} - \alpha)\tau) x^*(t - (\tfrac{1}{2} + \alpha)\tau) e^{-i2\pi\nu t} dt \right] \\ &= \int R_x(t + (\tfrac{1}{2} - \alpha)\tau, t - (\tfrac{1}{2} + \alpha)\tau) e^{-i2\pi\nu t} dt. \end{aligned}$$

Let  $R_x^{(\alpha)}(t, \tau) \equiv R_x(t + (\tfrac{1}{2} - \alpha)\tau, t - (\tfrac{1}{2} + \alpha)\tau)$ , then

$$EA_x^{(\alpha)}(\tau, \nu) = \int R_x^{(\alpha)}(t, \tau) e^{-i2\pi\nu t} dt. \quad (1)$$

If the EAF is zero about a given "TF lag point"  $(\tau_{12}, \nu_{12})$ , then any two TF points  $(t_1, f_1)$  and  $(t_1, f_2)$  with  $t_1 - t_2 = \tau_{12}$  and  $f_1 - f_2 = \nu_{12}$  are uncorrelated. The EAF indicates the *potential correlation* between Time-Frequency points separated by the time lag  $\tau$  and the frequency lag  $\nu$  [3].

Let  $[-\sigma_\tau, \sigma_\tau] \times [-\sigma_\nu, \sigma_\nu]$  be the smallest rectangle (centered at the origin of the  $(\tau, \nu)$  plane) which contains the effective support of the EAF, i.e.,

$$|EA_x^{(\alpha)}(\tau, \nu)| \approx 0 \quad \text{for } |\tau| > \sigma_\tau \text{ or } |\nu| > \sigma_\nu.$$

Define the **spread** of  $x$  as the area of this rectangle:

$$\sigma_x = 4\sigma_\tau\sigma_\nu. \quad (2)$$

From these definitions we can view the spread as a product of temporal correlation,  $\sigma_\tau$ , and spectral correlation  $\sigma_\nu$ .

It was shown that the EAFs obtained for different choices of  $\alpha$  are equal up to a phase factor [3],

$$EA_x^{(\alpha_2)}(\tau, \nu) = e^{i2\pi(\alpha_1 - \alpha_2)\tau\nu} EA_x^{(\alpha_1)}(\tau, \nu) \quad (3)$$

$$|EA_x^{(\alpha_2)}(\tau, \nu)| = |EA_x^{(\alpha_1)}(\tau, \nu)|. \quad (4)$$

Thus, with respect to the calculation of spread, the factor  $\alpha$  is arbitrary.

### 1.3. Numerical Spread

To define numerical spread we let  $T$  represent the sampling rate and denote

$$r_x[n, m] = R_x(nT, mT). \quad (5)$$

We can then write equation (1) in terms of discrete variables. Let

$$r_x^{(\alpha)}[n, m] = r_x[n + (\alpha - \tfrac{1}{2})m, n - (\alpha + \tfrac{1}{2})m] \quad (6)$$

then the numerical spread is given by

$$NEA_x^{(\alpha)}[m, k] = \sum_{n=0}^{N-1} r_x^{(\alpha)}[n, m] e^{-i2\pi nk/N}.$$

It was shown in Equation (4) that the choice of  $\alpha$  was arbitrary with respect to the calculation of spread. For discrete data we must choose  $\alpha = k/2$  for some integer  $k$  since this corresponds to integer shifts between  $n$  and  $m$  in equation (6). For our work, we chose  $\alpha = 1/2$  since it fits the above criterion and has the added benefit that the discrete version of the EAF, denoted NEA, becomes the discrete Fourier transform along the diagonals of the covariance matrix. We can now define the NEAF as,

$$NEA_x[m, k] = \sum_{n=0}^{N-1} r_x^{(1/2)}[n, m] e^{-i2\pi nk/N} \quad (7)$$

$$= \sum_{n=0}^{N-1} \hat{r}_x[n, n-m] e^{-i2\pi nk/N}, \quad (8)$$

where the  $\hat{r}_x$  is the extended autocovariance function which is the focus of this research and will be developed in the next section.

To calculate the spread we must determine the effective region of support of the NEA. To facilitate this calculation we project the signal onto the  $\tau$  and  $\nu$  axes and determine the support of these projections. Define the two projections as:

$$\tilde{N}_\tau[k] \equiv \sum_m NEA_x[m, k],$$

$$\tilde{N}_\nu[m] \equiv \sum_k NEA_x[m, k].$$

Using these projections we can calculate the spread in the  $\tau$  and  $\nu$  directions.

$$\sigma_\tau = |MT| : \{ |\tilde{N}_\nu[m]| < \delta, \text{ for } |m| > |M| \}, \quad (9)$$

$$\sigma_\nu = |\frac{K}{NT}| : \{ |\tilde{N}_\tau[k]| < \delta, \text{ for } |k| > |K| \}, \quad (10)$$

where  $\delta$  is a pre-determined threshold.

The **numerical spread** is defined as the effective support of the NEAF [1]

$$\sigma = 4\sigma_\tau\sigma_\nu.$$

## 2. THE EXTENDED AUTOCOVARIANCE

The extended autocovariance in Equation (8) arises from the finite nature of the data. If we assume that the data is a square covariance matrix it follows that the number of elements in a diagonal entry is a function of  $\tau$ ; if  $R \in \mathbb{R}^{N \times N}$  then the lengths of the diagonal

vectors will range in length from 1 to  $N$ . In order to maintain a constant frequency resolution in the  $\tau \times \nu$  plane, the diagonals must be padded with values such that each vector has length  $N$ .

The parameter  $\sigma_\nu$  can be viewed as the “effective bandwidth” of the covariance. The goal of the extension is to provide constant resolution with respect to  $\nu$  in the  $\tau \times \nu$  plane. However, by extending the data to achieve constant resolution, we run the risk of introducing artifacts into the NEAF that may ultimately effect the numerical spread.

Two issues must be addressed:

1. an increase in the energy of the diagonal entry could inflate the value of  $\sigma_\tau$ .
2. sharp discontinuities will add high frequency terms to the NEAF possibly inflating the value of  $\sigma_\nu$ .

Issue 1 can easily be addressed by scaling the extended covariance data so the energy along each diagonal is preserved. To address issue 2 three extension methods were examined. The extension methods chosen were based in part, on schemes examined by Karlsson and Vetterli [5] in the context of multirate filter banks.

We extended the computational matrix by increasing the range of  $m$  with  $-(N-1) \leq m \leq 2(N-1)$  and implemented the following schemes: zero padding, circular extension, and replication of edge values. Conceptually these are all straightforward methods of data manipulation. The method of zero padding simply pads the vector with zeros so that the resulting vector has length  $N$ . The method of circular extension creates the length  $N$  vector by treating the defined data as one period of a periodic sequence and replicating that sequence throughout the vector. The last method, replication of edge values, simply repeats the last defined value to form a vector of length  $N$ .

To be precise, we present these methods formally. To help simplify the presentation we introduce the auxiliary variable  $p = n - m$ .

1. Zero padding:

$$\hat{r}_x[n, m] = \begin{cases} r_x[n, m] & 0 \leq n, m \leq N-1, \\ 0 & \text{elsewhere.} \end{cases} \quad (11)$$

2. Circular extension:

$$\hat{r}_x[n, m] = r_x[\hat{n}, \hat{m}],$$

where

$$\hat{m} = \begin{cases} [(m - |p|) \bmod (N - |p|)] + |p|, & p < 0, \\ m \bmod (N - |p|), & p > 0, \\ m & p = 0, \end{cases} \quad (12)$$

and

$$\hat{n} = \begin{cases} n \bmod (N - |p|), & p < 0, \\ [(n - |p|) \bmod (N - |p|)] + |p|, & p > 0, \\ n & p = 0. \end{cases} \quad (13)$$

3. Replication of edge values: recall that  $m$  is defined such that  $-(N-1) \leq m \leq 2(N-1)$ , then

$$\hat{r}_{n,m} = \begin{cases} r_x[|p|, 0] & m < 0, \\ r_x[n, m] & 0 \leq m \leq N-1, \\ r_x[N-1-|p|, N-1] & m > N-1. \end{cases} \quad (14)$$

It should be noted that the other methods mentioned by Karlsson *et al.*, symmetric extension and doubly symmetric extension, are not well defined when applied to a data set with a defined length significantly shorter than the desired data length and so were not implemented here.

### 3. EXPERIMENTAL RESULTS

To compare the three extension schemes developed above they were implemented in Matlab and applied to multiple data sets. The results presented here arose from applying these methods to a covariance that slowly decayed away from the main diagonal and varied in time as shown in Figure 1. The size of the matrix was  $256 \times 256$ , the threshold used to determine the spread was  $\delta = 0.001$ .

#### 3.1. Zero Padding

The method of zero padding is easy to implement and automatically preserves energy but introduces jump discontinuities for any diagonal with a non-zero final entry. Jump discontinuities lead to high frequencies terms in the FFT and tend to inflate the value of  $\sigma_\nu$ . The resulting autocovariance and contour plot of the NEA are shown in Figure 2. The resulting numerical spread was  $\sigma = 940.637$ .

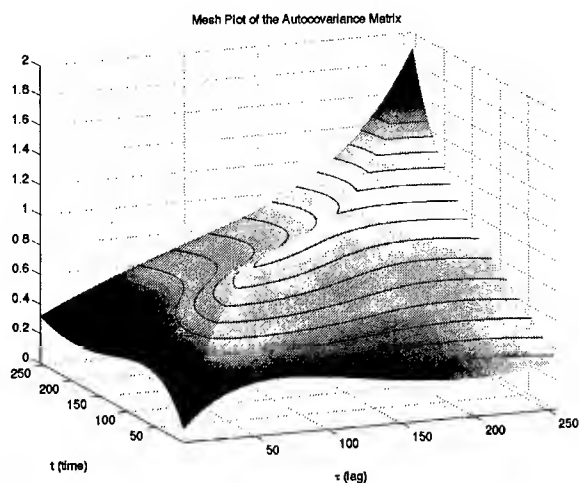


Figure 1: The covariance used as the test data shows variation in time and lag.

### 3.2. Circular Extension

The method of circular extension was applied to the same covariance data. In order to prevent inflation of  $\sigma_\tau$  each diagonal is normalized to preserve energy. Circular extension introduces jump discontinuities whenever the first and last data point on a diagonal differ, thus possibly inflating the value of  $\sigma_\nu$ . The resulting numerical spread was  $\sigma = 564.3822$ .

### 3.3. Replication of Edge Values

In Figure 4 we can see the resulting autocovariance and NEA when the method of edge value replication is used. This method produced the sharpest estimate of the support of the NEA for this data set. By replicating the edge values jump discontinuities are not induced. To prevent inflation of  $\sigma_\tau$  each diagonal was normalized as in the case of circular extension. The resulting numerical spread was  $\sigma = 514.2149$ .

Method	$\sigma_\nu$	$\sigma_\tau$	$\sigma$
Zero Padding	2.2457	214.5	1926.9
Circular Extension	2.2212	185.0	1648.1
Edge Replication	1.6812	176.0	1183.6

Table 1: Comparison of spread estimates by extension method with  $\delta = 0.001$ .

## 4. SUMMARY AND CONCLUSIONS

In this paper we have refined the technique of computing the numerical spread as introduced in earlier work.

We implemented three different schemes to augment the covariance matrix in order to calculate the Numerical Expected Ambiguity function on a grid of constant resolution. Of the three methods, zero padding, circular extension, and replication of edge values, it was determined that replication of edge values was least likely to inflate estimations of the frequency spread  $\sigma_\nu$ .

One theoretical justification for the improved performance of replication of edge values is that if  $x$  is stationary the diagonals of the correlation matrix will be constant *i.e.*,  $C$  is Toeplitz. By extending the data with constant values along the diagonals were are not artificially increasing the non-stationarity of the data which could increase the spread.

In future work we will examine the issue of robustness of numerical spread and suggest modifications for improved robustness. Thus providing both an enhanced framework for our future work in the area of adaptive signal representations, including the adaptation of novel multirate and wavelet signal processing techniques [6, 7].

## REFERENCES

- [1] R. A. Hedges and B. W. Suter, "The introduction of numerical spread as an indicator of non-stationarity," in *32nd Symposium on the Interface of Computing Science and Statistics*, 2000, vol. 33, accepted.
- [2] D.L. Donoho, S. Mallat, and R. von Sachs, "Estimating covariances of locally stationary processes: Rates of convergences of best basis methods," Tech. Rep., Stanford University, February 1998.
- [3] W. Kozek, F. Hlawatsch, H. Kirchauer, and U. Trautwein, "Correlative time-frequency analysis and classification of nonstationary random processes," in *Proceedings of the IEEE-SP International Symposium on Time-Frequency and Time-Scale Analysis*, 1994, pp. 417-420.
- [4] L. Cohen, *Time-Frequency Analysis*, Prentice Hall Signal Processing Series. Prentice Hall PTR, 1995.
- [5] Gunnar Karlsson and Martin Vetterli, "Extension of finite length signals for sub-band coding," *Signal Processing*, vol. 17, pp. 161-168, 1989.
- [6] B. W. Suter, *Multirate and Wavelet Signal Processing*, vol. 8 of *Wavelets and its Applications*, Academic Press, 1998.
- [7] R. A. Hedges, *Hybrid Wavelet Packet Analysis: Theory and Implementations*, Ph.D. thesis, Arizona State University, 1999.

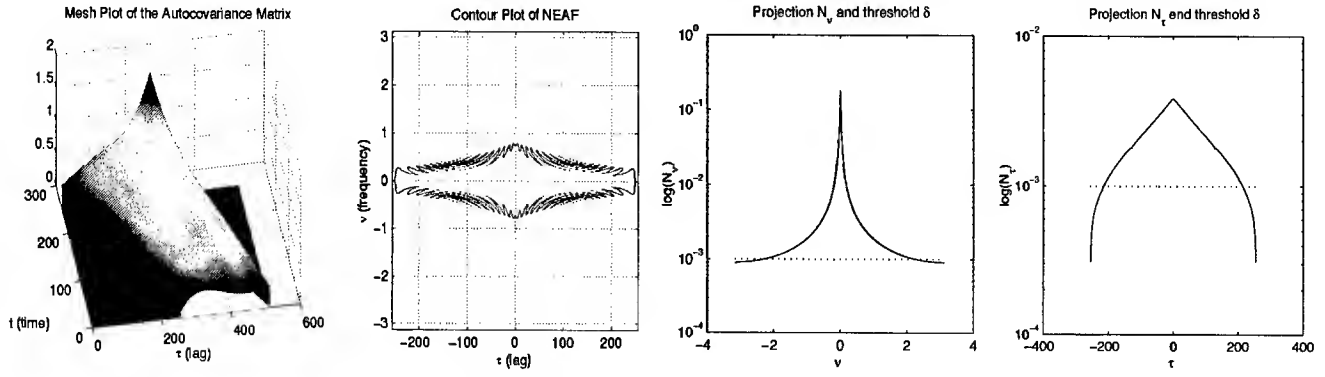
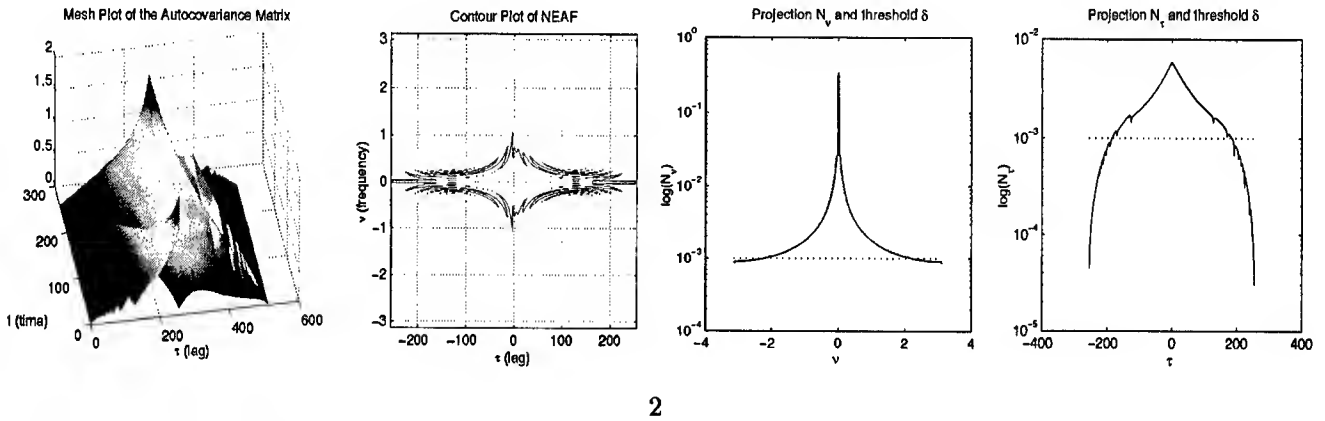


Figure 2: Edge Replication: mesh plot of  $\mathbf{R}_x^{(1/2)}[n, m]$ , a contour of the associated NEA, the projections  $\tilde{N}_\tau$  and  $\tilde{N}_\nu$  with threshold  $\delta = 0.001$  which yield  $\sigma_\tau = 214.5.0$  and  $\sigma_\nu = 2.2457$  thus  $\sigma = 1926.9$ .



2

Figure 3: Edge Replication: mesh plot of  $\mathbf{R}_x^{(1/2)}[n, m]$ , a contour of the associated NEA, the projections  $\tilde{N}_\tau$  and  $\tilde{N}_\nu$  with threshold  $\delta = 0.001$  which yield  $\sigma_\tau = 185.0$  and  $\sigma_\nu = 2.2212$  thus  $\sigma = 1648.1$ .

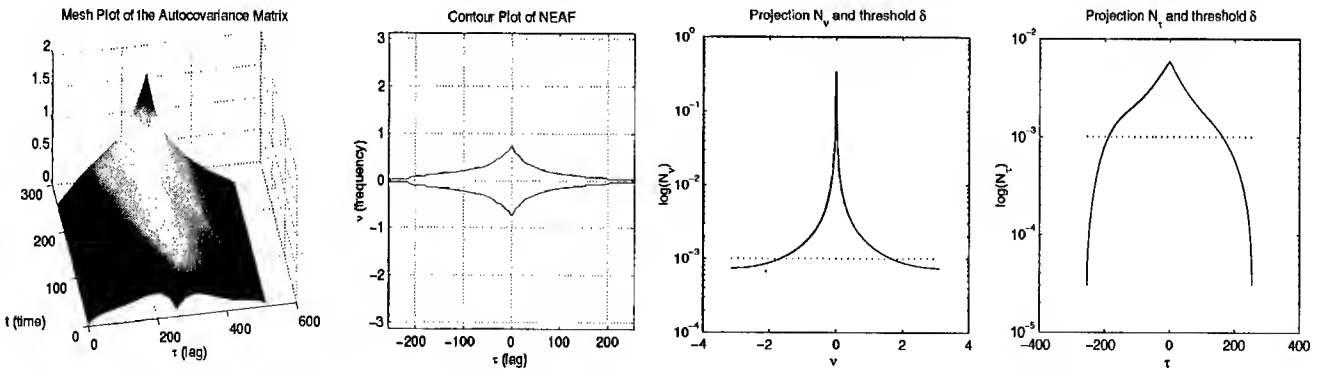


Figure 4: Edge Replication: mesh plot of  $\mathbf{R}_x^{(1/2)}[n, m]$ , a contour of the associated NEA, the projections  $\tilde{N}_\tau$  and  $\tilde{N}_\nu$  with threshold  $\delta = 0.001$  which yield  $\sigma_\tau = 176.0$  and  $\sigma_\nu = 1.6812$  thus  $\sigma = 1183.6$ .

# LOCALLY STATIONARY PROCESSES

Mark E. Oxley and Thomas F. Reid

Department of Mathematics and Statistics  
Air Force Institute of Technology  
2950 P Street  
Wright-Patterson AFB, OH 45433-7765  
Phone: (937)255-3636, FAX (937)656-4413  
Mark.Oxley@afit.af.mil, ext 4515  
Thomas.Reid@afit.af.mil, ext 4516

Bruce W. Suter

Air Force Research Laboratory  
Information Directorate  
525 Brooks Rd  
Rome, NY 13441-4505  
Phone: (315)330-7563, FAX (315)330-3908  
suterb@rl.af.mil

## ABSTRACT

The signals that arise in Air Force applications typically has noise that can be modeled as a non-stationary stochastic process. But, there may be intervals of time where the noise behaves more like a stationary process. This motivates the study of locally stationary stochastic processes. We rigorously define locally stationary stochastic processes and present their properties and relationships to stationary processes.

**Keywords:** stochastic process, stationary, piecewise stationary, locally stationary

## 1. INTRODUCTION

Properties of stationary processes are well known and have been used extensively in analyzing system performance and finding optimal controls for stochastic systems. Recall that a stationary process is one where all the finite-dimensional joint distributions are invariant to shifts in time (time homogeneous). Non-stationary processes appear in many engineering applications where random fluctuations change in time or space. If the process is slowly varying and if the interval is short enough, then the process can be approximated (in some sense) by a stationary one.

Recently, researchers (e.g., Mallat et. al.[5], Donoho [2] and Dahlhaus[1]) have turned their attention to so-called *locally stationary* processes as a tool to model systems where the behavior varies as a function of time. Unfortunately, to date there has not been an universality satisfying definition of what is meant by "locally stationary." This paper proposes such a definition, and illustrates the definition with some properties.

An early paper by Silverman [6] uses the term *locally stationary* to refer to a process whose covariance is a product of a (normalized) stationary covariance

multiplied by a sliding power factor. It appears that the "local" refers to a point property (verses an interval property). The efforts by Mallat et.al.[5] and Dahlhaus [1] have achieved some limited success in formalizing a definition of locally stationarity, in part, because theirs are based on Fourier analysis. Mallat uses the local trigonometric bases which originated with Coifman and Meyer [?] and has been generalized by Suter and Oxley [7]. On the other hand, Dahlhaus' definition of locally stationary allows a much broader class of processes, including processes where the mean is never constant for any given time period. There are other references in the literature working with locally stationary processes, e.g. [3], but they do not define locally stationary. It appears that their working definition is similar to Mallat's.

Donoho et. al. [2] use a particular definition of *locally stationary*, tailored to allow them to study certain phenomenon of time-inhomogeneity. Their definition allows very abrupt changes in the process (allowing a locally stationary window to be as short as one sample) so long as the correlation between samples decays sufficiently fast and there are not "too many" change points in a given interval. In the end, they call for the development of a new definition since "there is, at the moment, no definition which really captures all the facets of local stationarity."

## 2. DEFINITIONS

Let  $(\Omega, \mathcal{B}, P)$  be a probability space, so that  $\Omega$  is the sample set of outcomes,  $\mathcal{B}$  is the  $\sigma$ -algebra of events, and  $P$  is the probability measure. Thus, a real-valued random variable  $X$  is a function mapping an outcome  $\omega \in \Omega$  to some real number  $r \in \mathbb{R}$ , so that  $r = X(\omega)$ . Let  $I \subset \mathbb{R}$  be an interval (possibly the whole real

line). For each  $t \in I$ , let  $X(t)$  be a random variable, so that  $X(t, \omega)$  is a real number. A real-valued stochastic process  $\mathbf{X} = \{X(t) : t \in I\}$  is a collection of random variables. Let  $S \subset I$  be a finite set, i.e.,  $S = \{t_1, t_2, \dots, t_n\}$  with  $t_i \neq t_j$  for  $i \neq j$  for some  $n \in \mathbb{N}$ . Denote the collection of all possible finite subsets of  $I$  by  $\mathcal{F}_I$ . For a real number  $r$  we use the notation  $S + r = \{s_1 + r, s_2 + r, \dots, s_n + r\}$  to denote the translated set. Given  $S = \{t_1, t_2, \dots, t_n\} \in \mathcal{F}_I$ , define the ordered set of random variables  $\mathcal{X}(S) = (X(t_1), X(t_2), \dots, X(t_n))$ . Define the joint cumulative distribution function of these random variables to be

$$F_{\mathcal{X}(S)}(\mathbf{c}) = \Pr[X(t_1) \leq c_1, X(t_2) \leq c_2, \dots, X(t_n) \leq c_n]$$

where  $\mathbf{c} = (c_1, c_2, \dots, c_n) \in \mathbb{R}^n$ . Recall the definition of a stationary stochastic process.

**Definition 1** (*Stationary on the real numbers*). A stochastic process  $\{X(t) : t \in \mathbb{R}\}$  is said to be stationary on the real numbers if

$$F_{\mathcal{X}(S)} = F_{\mathcal{X}(S+r)}$$

for all  $r \in \mathbb{R}$ , for all  $S \in \mathcal{F}_{\mathbb{R}}$ .

This definition of stationary is also called *strictly stationary*. If the cardinality of  $S$  is two, this definition becomes *wide-sense stationary*. Many real-world systems may not meet the strict requirements for stationarity, but, if we consider an interval,  $I$ , then it would be stationary with respect to shifts within that interval. This leads to the following definition of stationary on an interval.

**Definition 2** (*Stationary on an interval*). Let  $I \subset \mathbb{R}$  be an interval. A stochastic process  $\{X(t) : t \in I\}$  is said to be stationary on the interval  $I$  if

$$F_{\mathcal{X}(S)} = F_{\mathcal{X}(S+r)}$$

for all  $r \in \delta(S, I)$ , for all  $S \in \mathcal{F}_I$ , where  $\delta(S, I) = \{r \in \mathbb{R} : S + r \subset I\}$ .

Notice that the interval  $I$  could be the real numbers, in such case this definition reduces to strictly stationary. Thus, this definition is more general.

Recall the definition of a partition of an interval.

**Definition 3** (*Partition*). Let  $I \subset \mathbb{R}$  be an interval (possibly  $\mathbb{R}$ ). A partition of  $I$  is a countable collection of subintervals  $\{J_1, J_2, \dots\}$  where  $J_k \subset I$  is an interval for each  $k \in \mathcal{I}$ , some countable index set, such that

1.  $J_i \cap J_k = \emptyset$ , (the empty set) for all  $i \neq k$  in  $\mathcal{I}$ .

$$2. \bigcup_{k \in \mathcal{I}} J_k = I.$$

Recall that countable includes finite, thus the index set  $\mathcal{I}$  may be the finite counting set  $\mathcal{I} = \{1, 2, \dots, K\}$  for some positive integer  $K$  and the partition is  $\{J_1, J_2, \dots, J_K\}$ . We will denote a partition of the interval  $I$  by  $P$ , and we let  $\mathfrak{P}_I$  represent the collection of all possible partitions of the interval  $I$ . Thus  $P = \{J_1, J_2, \dots, J_K\} \in \mathfrak{P}_I$  is a specific instantiation of the collection.

**Definition 4** (*Locally stationary on an interval*). Let  $I \subset \mathbb{R}$  be an interval (possibly  $\mathbb{R}$ ). A stochastic process  $\{X(t) : t \in I\}$  is said to be locally stationary on  $I$  if there exists some partition  $P \in \mathfrak{P}_I$  and at least one subinterval  $J \in P$  such that the stochastic process  $\{X(t) : t \in J\}$  is stationary on  $J$ .

**Definition 5** (*Piecewise stationary on an interval*). Let  $I \subset \mathbb{R}$  be an interval (possibly  $\mathbb{R}$ ). A stochastic process  $\{X(t) : t \in I\}$  is said to be piecewise stationary on  $I$  if there exists some partition  $P \in \mathfrak{P}_I$  such that on all subintervals  $J \in P$  the stochastic process  $\{X(t) : t \in J\}$  is stationary on  $J$ .

This last definition is the continuous version corresponding to the discrete version used by Levielle [4], where the determination of the change point was sought.

### 3. PROPERTIES OF LOCALLY STATIONARY PROCESSES.

The section investigates some of the properties that follow from our definitions. To remove any confusion with the locally stationary process, we will call a process that is stationary on the (whole) interval a *globally stationary process*. Now, we have three types of stationary processes under consideration: globally, locally and piecewise. In this section we give properties of each type and also their relations to each other. First, we establish some notation to clarify the presentation. All processes are assumed to be real-valued. The extension to complex-valued processes is obvious.

Let  $\mathcal{S}$  denote the set of stochastic processes defined on  $I$ ,  $\mathcal{G}$  denote the set of globally stationary processes defined on  $I$ ,  $\mathcal{L}$  denote the set of locally stationary processes defined on  $I$ , and  $\mathcal{P}$  denote the set of piecewise stationary processes defined on  $I$ . Therefore,

$$\begin{aligned} \mathcal{S} &= \{\mathbf{X} \text{ is stochastic process on } I\} \\ \mathcal{G} &= \{\mathbf{X} \text{ is globally stationary on } I\} \\ \mathcal{L} &= \{\mathbf{X} \text{ is locally stationary on } I\} \\ \mathcal{P} &= \{\mathbf{X} \text{ is piecewise stationary on } I\}. \end{aligned}$$

Define stochastic process equality  $\stackrel{S}{=}$  to be pointwise equality. Thus, for  $\mathbf{X}, \mathbf{Y} \in \mathcal{S}$

$$\mathbf{X} \stackrel{S}{=} \mathbf{Y} \text{ if and only if } X(t, \omega) \stackrel{\mathbb{R}}{=} Y(t, \omega)$$

for all  $t \in I, \omega \in \Omega$ . Here  $\stackrel{\mathbb{R}}{=}$  denotes real equality.

**Theorem 1**  $\mathcal{G} \subset \mathcal{P} \subset \mathcal{L} \subset \mathcal{S}$ .

Define stochastic process addition  $\stackrel{S}{+}$  for  $\mathbf{X}, \mathbf{Y} \in \mathcal{S}$  to be

$$[\mathbf{X} \stackrel{S}{+} \mathbf{Y}](t, \omega) \stackrel{\mathbb{R}}{=} X(t, \omega) + Y(t, \omega)$$

for all  $t \in I, \omega \in \Omega$ . Here  $+$  denotes real addition.

Define the zero stochastic process  $\mathbf{Z}$  to be  $Z(t, \omega) \stackrel{\mathbb{R}}{=} 0$  for all  $t \in I, \omega \in \Omega$ . ( $\mathbf{Z}$  is the identity with respect to  $\stackrel{S}{+}$ .) There are several algebraic properties concerning these sets.

**Theorem 2**  $\mathcal{S}, \mathcal{G}$  and  $\mathcal{P}$  are closed with respect to  $\stackrel{S}{+}$  addition.  $\mathcal{L}$  is not closed with respect to  $\stackrel{S}{+}$  addition.

Therefore, if  $\mathbf{X}, \mathbf{Y} \in \mathcal{P}$  then  $\mathbf{X} \stackrel{S}{+} \mathbf{Y} \in \mathcal{P}$ . If  $\mathbf{X}, \mathbf{Y} \in \mathcal{L}$  then  $\mathbf{X} \stackrel{S}{+} \mathbf{Y}$  may or may not be locally stationary.

Define stochastic process multiplication  $\stackrel{S}{\cdot}$  for  $\mathbf{X}, \mathbf{Y} \in \mathcal{S}$  to be pointwise multiplication,

$$[\mathbf{X} \stackrel{S}{\cdot} \mathbf{Y}](t, \omega) \stackrel{\mathbb{R}}{=} X(t, \omega) \cdot Y(t, \omega)$$

for all  $t \in I, \omega \in \Omega$ . Here  $\cdot$  denotes real multiplication.

Define the unit or identity stochastic process  $\mathbf{U}$  to be  $U(t, \omega) \stackrel{\mathbb{R}}{=} 1$  for all  $t \in I, \omega \in \Omega$ . ( $\mathbf{U}$  is the identity with respect to  $\stackrel{S}{\cdot}$ .)

**Theorem 3**  $\mathcal{S}, \mathcal{G}$  and  $\mathcal{P}$  are closed with respect to  $\stackrel{S}{\cdot}$  multiplication.  $\mathcal{L}$  is not closed with respect to  $\stackrel{S}{\cdot}$  multiplication.

Therefore, if  $\mathbf{X}, \mathbf{Y} \in \mathcal{P}$  then  $\mathbf{X} \stackrel{S}{\cdot} \mathbf{Y} \in \mathcal{P}$ . If  $\mathbf{X}, \mathbf{Y} \in \mathcal{L}$  then  $\mathbf{X} \stackrel{S}{\cdot} \mathbf{Y}$  may or may not be locally stationary.

Define scalar multiplication  $\stackrel{sc}{\cdot}$  to be the following: given  $r \in \mathbb{R}$  and  $\mathbf{X} \in \mathcal{S}$

$$[r \stackrel{sc}{\cdot} \mathbf{X}](t, \omega) \stackrel{\mathbb{R}}{=} r \cdot X(t, \omega)$$

for all  $t \in I, \omega \in \Omega$ .

**Theorem 4**  $\mathcal{S}, \mathcal{G}, \mathcal{P}$  and  $\mathcal{L}$  are closed with respect to  $\stackrel{sc}{\cdot}$  multiplication.

Some pleasing algebraic results follow.

**Theorem 5**  $(\mathcal{S}, \stackrel{S}{=}, \stackrel{S}{+}, \stackrel{S}{\cdot})$ ,  $(\mathcal{G}, \stackrel{S}{=}, \stackrel{S}{+}, \stackrel{S}{\cdot})$  and  $(\mathcal{P}, \stackrel{S}{=}, \stackrel{S}{+}, \stackrel{S}{\cdot})$  are linear spaces over the  $\mathbb{R}$ . Furthermore,  $\mathcal{G}$  is a subspace of  $\mathcal{P}$  which is a subspace of  $\mathcal{S}$ .

**Theorem 6**  $(\mathcal{S}, \stackrel{S}{=}, \stackrel{S}{+}, \stackrel{S}{\cdot}, \stackrel{sc}{\cdot})$ ,  $(\mathcal{G}, \stackrel{S}{=}, \stackrel{S}{+}, \stackrel{S}{\cdot}, \stackrel{sc}{\cdot})$  and  $(\mathcal{P}, \stackrel{S}{=}, \stackrel{S}{+}, \stackrel{S}{\cdot}, \stackrel{sc}{\cdot})$  are commutative linear algebras with identity over the  $\mathbb{R}$ . Furthermore,  $\mathcal{G}$  is a subalgebra of  $\mathcal{P}$  which is a subalgebra of  $\mathcal{S}$ .

It is interesting that  $\mathcal{L}$  is "between"  $\mathcal{P}$  and  $\mathcal{S}$  but is not a linear space. If one requires some additional conditions then a linear space is attained.

**Definition 6** Let  $P$  be a partition of  $I$ , i.e.,  $P \in \mathfrak{P}_I$ . Let  $Q \subset P$  be a subpartition. Let  $\mathcal{L}_Q$  denote the collection of stochastic processes that are stationary on all subintervals  $J$  in the subpartition  $Q$ , i.e.,

$$\mathcal{L}_Q = \{\mathbf{X} \text{ is stationary on } J, \forall J \in Q\}.$$

**Theorem 7** Let  $P \in \mathfrak{P}_I$  and  $Q \subset P$  be a subpartition, then  $(\mathcal{L}_Q, \stackrel{S}{=}, \stackrel{S}{+}, \stackrel{S}{\cdot})$  is a linear subspace of  $(\mathcal{S}, \stackrel{S}{=}, \stackrel{S}{+}, \stackrel{S}{\cdot})$ .

**Definition 7** Let  $P \in \mathfrak{P}_I$ . Let  $\mathcal{P}_P$  denote the collection of piecewise stationary processes that are stationary on every subinterval of the partition  $P$ .

$$\mathcal{P}_P = \{\mathbf{X} \text{ is stationary on } J, \forall J \in P\}.$$

**Theorem 8** Let  $P \in \mathfrak{P}_I$  then  $(\mathcal{P}_P, \stackrel{S}{=}, \stackrel{S}{+}, \stackrel{S}{\cdot})$  is a linear space. If  $Q \subset P$  then  $(\mathcal{P}_P, \stackrel{S}{=}, \stackrel{S}{+}, \stackrel{S}{\cdot})$  is a subspace of  $(\mathcal{L}_Q, \stackrel{S}{=}, \stackrel{S}{+}, \stackrel{S}{\cdot})$ .

It is clear that if the subpartition is, in fact, the original partition, so that  $Q = P$ , then  $\mathcal{L}_Q = \mathcal{L}_P = \mathcal{P}_P$ . There are other interesting properties concerning  $\mathcal{L}_Q$ . One concerns the "mixing" of the subpartitions. If  $P_1$  and  $P_2$  are two different partitions in  $\mathfrak{P}_I$  and  $Q_1 \subset P_1$  and  $Q_2 \subset P_2$  are subpartitions then what can we say about "mixing" the subspaces  $\mathcal{L}_{Q_1}$  and  $\mathcal{L}_{Q_2}$ ? It will depend on the subpartitions.

**Theorem 9** Let  $P_1, P_2 \in \mathfrak{P}_I$ , and  $Q_1 \subset P_1, Q_2 \subset P_2$  be subpartitions. Let  $m$  denote the Lebesgue measure. If  $m(Q_1 \cap Q_2) = 0$  then  $\mathcal{L}_{Q_1 \cap Q_2} = \mathcal{S}$ . If  $m(Q_1 \cap Q_2) > 0$  then  $\mathcal{L}_{Q_1 \cap Q_2} = \mathcal{L}_{Q_1} \cap \mathcal{L}_{Q_2}$ .

In the case when the intersection  $Q_1 \cap Q_2$  has positive measure, then there exists a finer partition of both  $P_1$  and  $P_2$  (say  $P_3$ ) such that  $Q_1 \cap Q_2$  is a subpartition of  $P_3$ . When we union all these subspaces  $\mathcal{L}_Q$  we get  $\mathcal{L}$ .

**Theorem 10** The union over all the subspaces  $\mathcal{L}_Q$  yields  $\mathcal{L}$ , that is,

$$\mathcal{L} = \bigcup_{Q \subset P \in \mathfrak{P}_I} \mathcal{L}_Q.$$

Similarly for  $\mathcal{P}_P$ , that is,

$$\mathcal{P} = \bigcup_{P \in \mathfrak{P}_I} \mathcal{P}_P.$$

The next group of properties relate a stochastic process and a deterministic process, that is, a function.

**Theorem 11** If  $\mathbf{X} \in \mathcal{G}$  and  $f$  is piecewise constant function on  $I$ , then

1.  $Y(t) = f(t) + X(t)$  is piecewise stationary on  $I$ .
2.  $Y(t) = f(t) \cdot X(t)$  is piecewise stationary on  $I$ .
3.  $Y(t) = X(t + f(t))$  is piecewise stationary on  $I$  (only if  $I = \mathbb{R}$ ).

**Proof of 1.** Let  $J \subset I$  be a subinterval where  $f$  is constant, so that  $f(t) = k$  for  $t \in J$  and some constant  $k$ . Then, for any  $S \in \mathcal{F}_J$  and  $r \in \delta(S, J)$ , we have

$$\begin{aligned} F_{Y(S+r)} &= \Pr[f(t_1 + r) + X(t_1 + r) \leq c_1, \\ &\quad \dots, f(t_n + r) + X(t_n + r) \leq c_n] \\ &= \Pr[f(t_1) + X(t_1) \leq c_1, \\ &\quad \dots, f(t_n) + X(t_n) \leq c_n] \\ &= F_{Y(S)} \end{aligned}$$

since  $f(t_i + r) = f(t_i)$  for  $r \in \delta(S, J)$ .

Note that the process  $\{Y(t) : t \in J\}$  may have a mean which varies with time, but the covariance structure is the same as that of  $\{X(t) : t \in J\}$ . For  $s, t \in J$

$$\begin{aligned} \text{Cov}[Y(s), Y(t)] &= \text{E}[(f(s) - X(s) - \text{E}[f(s) - X(s)]) \\ &\quad \times (f(t) - X(t) - \text{E}[f(t) - X(t)])] \\ &= \text{E}[(X(s) - \text{E}[X(s)])(X(t) - \text{E}[X(t)])] \\ &= \text{Cov}[X(s), X(t)]. \end{aligned}$$

**Proof of 2.** Similar to 1. For this process, both the mean and the covariance structure may vary with time. For  $s, t \in J$

$$\begin{aligned} \text{Cov}[Y(s), Y(t)] &= \text{E}[(f(s)X(s) - \text{E}[f(s)X(s)]) \\ &\quad \times (f(t)X(t) - \text{E}[f(t)X(t)])] \\ &= f(s)f(t)\text{Cov}[X(s), X(t)]. \end{aligned}$$

**Proof of 3.** Similar to 1. For this case, the mean of  $\{Y(t) : t \in J\}$  is the same as for  $\{X(t) : t \in J\}$ , but

the covariance structure changes as a function of time. For  $s, t \in J$

$$\begin{aligned} \text{Cov}[Y(s), Y(t)] &= \text{Cov}[X(s + f(s)), X(t + f(t))] \\ &= \text{Cov}[X(s), X(t + f(t) - f(s))] \\ &= \text{Cov}[X(0), X(t - s + f(t) - f(s))]. \end{aligned}$$

**Theorem 12** Suppose the processes  $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_N \in \mathcal{G}$ , and the functions  $f_1, f_2, \dots, f_N$  are piecewise constant on  $I$ . Then the process  $\mathbf{Y}$  given by

$$Y(t) = \sum_{j=1}^N f_j(t) \cdot X_j(t)$$

is piecewise stationary on  $I$ .

For this process, both the mean and the covariance structure may vary with time, since for  $s, t \in J \in P \in \mathfrak{P}_I$

$$\begin{aligned} \text{Cov}[Y(s), Y(t)] &= \text{E}[(\sum_i f_i(s)X_i(s) - \text{E}[\sum_i f_i(s)X_i(s)]) \\ &\quad \times (\sum_j f_j(t)X_j(t) - \text{E}[\sum_j f_j(t)X_j(t)])] \\ &= \sum_i \sum_j f_i(s)f_j(t)\text{Cov}[X_i(s), X_j(t)]. \end{aligned}$$

**Definition 8** (Locally constant on an interval). Let  $I \subset \mathbb{R}$  be an interval (possibly  $\mathbb{R}$ ). A function  $f$  defined on  $I$  is said to be locally constant on  $I$  if there exists some partition  $P \in \mathfrak{P}_I$  and at least one subinterval  $J \in P$  such that  $f$  is constant on  $J$ . Let  $\mathcal{LC}$  denote the collection of locally constant functions defined on  $I$ , that is,

$$\mathcal{LC} = \{f : I \rightarrow \mathbb{R} \mid f \text{ is locally constant on } I\}.$$

**Theorem 13** If  $\mathbf{X} \in \mathcal{G}$  and  $f \in \mathcal{LC}$ , then

1.  $\mathbf{Y} = f + \mathbf{X} \in \mathcal{L}$ .
2.  $\mathbf{Y} = f \cdot \mathbf{X} \in \mathcal{L}$ .
3.  $Y(t) = X(t + f(t)) \in \mathcal{L}$  (only if  $I = \mathbb{R}$ ).

Notice that if we define the identity function  $i$  to be  $i(t) = t$  for all  $t \in I$ , then the last result is a composition of a stochastic process with a deterministic process (a function). That is,  $\mathbf{Y} = \mathbf{X} \circ (i + f)$ .

Other questions regarding composition remain to be asked. For example, what about  $g \circ \mathbf{X}$  processes? Let  $\mathbb{R}[x]$  denote the set of polynomials with real coefficients and indeterminant  $x$ , so that  $g \in \mathbb{R}[x]$  implies that  $g(x) = a_0 + a_1x + \dots + a_Nx^N$  for some degree  $N \in \mathbb{N}$  and coefficients  $a_0, a_1, \dots, a_N \in \mathbb{R}$ .

**Theorem 14** Let  $\mathbf{X} \in \mathcal{L}$ , then  $\mathbf{X}^2 \equiv \mathbf{X} \cdot \mathbf{X} \in \mathcal{L}$ . Furthermore, for each  $g \in \mathbb{R}[x]$  then  $g \circ \mathbf{X} \in \mathcal{L}$ .



**Theorem 15** For each  $\mathbf{X} \in \mathcal{L}$  the set  $\mathcal{L}_{\mathbf{X}} = \{\mathbf{Y} = g(\mathbf{X}) : g \in \mathbb{R}[x]\}$  is a commutative algebra with identity.

There are invariance results related to composition for the other types of stationarity.

**Theorem 16** Let  $g \in \mathbb{R}[x]$ .

1. If  $\mathbf{X} \in \mathcal{G}$  then  $g \circ \mathbf{X} \in \mathcal{G}$ .
2. If  $\mathbf{X} \in \mathcal{P}$  then  $g \circ \mathbf{X} \in \mathcal{P}$ .
3. If  $\mathbf{X} \in \mathcal{L}$  then  $g \circ \mathbf{X} \in \mathcal{L}$ .
4. If  $\mathbf{X} \in \mathcal{S}$  then  $g \circ \mathbf{X} \in \mathcal{S}$ .

#### 4. CONCLUSIONS

Recently Donoho [2] called for a new definition of locally stationarity since, "there is, at the moment, no definition which really captures all facets of local stationarity". We believe our definition *does* capture all such facets. Space does not allow for the other properties. More results are forthcoming concerning applications.

All these properties and theorems hold true if we replace the real numbers with complex numbers.

#### REFERENCES

- [1] R. Dahlhaus. Fitting time series models to nonstationary processes. *The Annals of Statistics*, 25:1–37, 1997.
- [2] D. Donoho, S. Mallat, and R. von Sachs. Estimating covariances of locally stationary processes: Rates of convergence of best basis methods. Technical Report 517, Stanford University, Dept of Statistics, February 1994.
- [3] C. L. Fancourt and J. C. Principe. Competitive principal component analysis for locally stationary time series. *IEEE Transactions on Signal Processing*, 46(11):3068–3081, November 1998.
- [4] M. Lavielle. Optimal segmentation of random processes. *IEEE Transactions on Signal Processing*, 46(5):1365–1373, May 1998.
- [5] S. Mallat, G. Papanicolaou, and Z. Zhang. Adaptive covariance estimation of locally stationary processes. *The Annals of Statistics*, 26:1–47, 1998.
- [6] R. A. Silverman. Locally stationary random processes. *IRE Transactions on Information Theory*, IT-3:182–187, 1957.
- [7] B. Suter and M. Oxley. On overlapped windows and weighted orthonormal bases. *IEEE Transactions on Signal Processing*, 42:1973–1982, 1994.

# STATISTICAL PERFORMANCE COMPARISON OF A PARAMETRIC AND A NON-PARAMETRIC METHOD FOR IF ESTIMATION OF RANDOM AMPLITUDE LINEAR FM SIGNALS IN ADDITIVE NOISE

Mark R. Morelande, Braham Barkat and Abdelhak M. Zoubir

Australian Telecommunications Research Institute  
School of Electrical and Computer Engineering  
Curtin University of Technology, GPO Box U1987, Perth WA 6845, Australia  
email: mark@atri.curtin.edu.au, rbarkatb@curtin.edu.au

## ABSTRACT

This paper presents a statistical performance comparison between the cyclic moments-based and Wigner-Ville distribution-based instantaneous frequency estimators for linear FM signals in real-valued multiplicative and complex-valued additive noise. Theoretical results are used to compare the performance of the estimation algorithms over a wide range of conditions. Simulation results confirm our theoretical derivations.

## 1. INTRODUCTION

Accurate estimation of the instantaneous frequency (IF) of frequency modulated (FM) signals is important in many engineering applications such as radar, sonar, acoustic emission and telecommunications [2]. In many cases it is assumed that the signal of interest has a constant amplitude. While this is a valid assumption in a wide range of scenarios, there are several important applications in which the constant amplitude assumption is inappropriate. Examples include fading in wireless communications [13] and the case of a fluctuating target in radar [14]. In both of these cases the amplitude varies randomly with time and the discrete-time signal may be written as

$$X_t = A_t s_t + W_t, \quad t = 0, \dots, n-1 \quad (1)$$

where  $s_t = \exp(j\phi_t)$ ,  $A_t$  is a real-valued stationary random process with mean  $\mu_A$  and variance  $\sigma_A^2$  and  $W_t$  is a complex-valued stationary random process, independent of  $A_t$ , with variance  $\sigma_W^2$ .

The methods available for estimating the IF of  $s_t$  may be separated into two classes. One class of estimators consists of parametric methods in which the phase  $\phi_t$  is modelled by a sum of basis sequences,

$$\phi_t = \sum_{i=0}^q b_i \gamma_{t,i} \quad (2)$$

An IF estimate can be obtained from estimates of the phase parameters  $b_0, \dots, b_q$ . A commonly used set of basis sequences is  $\gamma_{t,i} = t^i$ ,  $i = 0, \dots, q$ , in which case the signal  $s_t$  is referred to as a polynomial phase signal (PPS). Estimators of the phase parameters for the PPS model have been proposed in [4, 12].

The second class of IF estimators consists of non-parametric methods. A subclass of these non-parametric methods are based on time-frequency distributions such as the Wigner-Ville distribution (WVD) and its higher-order generalisation, the polynomial WVD

(PWVD) [3]. These estimators may be applied regardless of the form of  $\phi_t$  though it should be kept in mind that a systematic bias will result unless  $s_t$  is a PPS. If  $s_t$  is not a PPS accurate IF estimates can still be obtained from the PWVD by using an adaptive window [1].

The purpose of this paper is to compare the performances of the cyclic moments-based procedure [12] and the WVD-based procedure for estimating the IF of  $s_t$  from observations  $x_0, \dots, x_{n-1}$  of (1). The signal  $s_t$  is assumed to be a unit modulus PPS of order two i.e. the phase  $\phi_t$  is given by (2) with  $q = 2$  and  $\gamma_{t,i} = t^i$ ,  $i = 0, 1, 2$ ,  $t = 0, \dots, n-1$ . Theoretical expressions for the variances of both estimators are given and compared for various noise scenarios. The theoretical results are confirmed using numerical simulations.

The paper is structured as follows. Section 2 contains a brief review of the WVD-based and cyclic moments-based IF estimators. In this section we also verify the suitability of the WVD for IF estimation of random amplitude linear FM signals. Expressions for the variances of these estimators are given in section 3. The theoretical results are used to compare the WVD-based and cyclic moments-based IF estimators in section 4. Simulation results are included to confirm the theoretical results. The paper concludes with a discussion of the main results.

## 2. REVIEW OF ESTIMATORS

In this section we review the WVD-based and cyclic moments-based IF estimators. We also present new results which verify the suitability of the WVD for IF estimation of random amplitude linear FM signals.

### 2.1. WVD-based IF Estimator

It is well-known that the peak of the WVD can be used to estimate the IF of constant amplitude linear FM signals [11]. In this paper we propose the use of the WVD for estimating the IF of random amplitude linear FM signals. We show that, contrary to statements made in [3] regarding the necessity of higher-order PWVDs, the WVD may be used to estimate the IF of random amplitude linear FM signals.

The discrete-time WVD of the sequence  $x_0, \dots, x_{n-1}$  is de-

defined as

$$W_{2X}(t, \omega) = \frac{1}{2m+1} \sum_{\xi=-m}^m x_{t+\xi} x_{t-\xi}^* \exp(-j2\omega\xi) \quad (3)$$

where  $m = \min(n-1-t, t)$ . Throughout the paper we use upper-case letters e.g.  $X_t$  to denote random variables and lower-case letters e.g.  $x_t$  to denote realisations of random variables. Note that scaling by the window length  $1/(2m+1)$  is introduced for the sake of normalisation and does not affect the statistical properties of the WVD-based IF estimator. Substituting for  $x_t$  gives

$$W_{2X}(t, \omega) = \frac{1}{2m+1} \sum_{\xi=-m}^m y_{t+\xi} y_{t-\xi}^* \exp\{-j2(\omega - \omega_t)\xi\} \quad (4)$$

where  $Y_t = A_t + U_t$ ,  $U_t = W_t/s_t$  and  $\omega_t = b_1 + 2b_2t$  is the IF of  $s_t$ . Assuming that the additive noise is complex white Gaussian we obtain

$$\begin{aligned} E W_{2X}(t, \omega) &= \frac{1}{2m+1} \sum_{\xi=-m}^m \{\mu_A^2 + c_{2A}(2\xi) + \sigma_W^2 \delta_\xi\} \\ &\times \exp\{-j2(\omega - \omega_t)\xi\} \quad (5) \\ &= \frac{1}{2m+1} \{\mu_A^2 \Delta^{(2m+1)}(\omega - \omega_t) \\ &+ C_{2A}(\omega) * \Delta^{(2m+1)}(\omega - \omega_t) + \sigma_W^2\} \quad (6) \end{aligned}$$

where  $\delta_\xi$  is the Kronecker delta,  $c_{2A}(\xi) = \text{cum}(A_t, A_{t+\xi})$  is the second-order cumulant or covariance of  $A_t$ ,

$$C_{2A}(\omega) = \sum_{\xi=-\infty}^{\infty} c_{2A}(\xi) \exp(-j\omega\xi)$$

is the second-order spectrum of  $A_t$ , '\*' denotes the convolution operation and

$$\Delta^{(l)}(\omega) = \sin(l\omega)/\sin(\omega). \quad (7)$$

Using the additivity property of the cumulant [9], the variance of the WVD may be written as

$$\begin{aligned} \text{var}\{W_{2X}(t, \omega)\} &= \frac{1}{(2m+1)^2} \sum_{\xi_1=-m}^m \sum_{\xi_2=-m}^m \\ &\text{cum}(Y_{t+\xi_1} Y_{t-\xi_1}^*, Y_{t+\xi_2} Y_{t-\xi_2}^*) \exp\{-j2(\omega - \omega_t)(\xi_1 + \xi_2)\} \quad (8) \end{aligned}$$

For large window lengths we obtain,

$$\begin{aligned} \text{var}\{W_{2X}(t, \omega)\} &= \frac{1}{(2m+1)^2} \sum_{\xi_1=-m}^m \sum_{\xi_2=-m}^m \\ &[c_{2A}(\xi_1 - \xi_2)^2 + \{c_{2A}(\xi_1 + \xi_2) + \sigma_W^2 \delta_{\xi_1 + \xi_2}\}^2 \\ &+ 2\mu_A^2 \{c_{2A}(\xi_1 - \xi_2) + c_{2A}(\xi_1 + \xi_2) + \sigma_W^2 \delta_{\xi_1 + \xi_2}\}] \\ &\times \exp\{-j2(\omega - \omega_t)(\xi_1 + \xi_2)\} + O(m^{-2}), \quad (9) \end{aligned}$$

where we have assumed the mixing condition,

$$\sum_{\xi_1, \dots, \xi_{p-1}=-\infty}^{\infty} |c_{pA}(\xi_1, \dots, \xi_{p-1})| < \infty, \quad p = 1, \dots, 4 \quad (10)$$

where  $c_{pA}(\xi_1, \dots, \xi_{p-1}) = \text{cum}(A_t, A_{t+\xi_1}, \dots, A_{t+\xi_{p-1}})$ . We see from (9) and (10) that  $\text{var}\{W_{2X}(t, \omega)\}$  is  $O(m^{-1})$ . It follows from the Chebyshev inequality that, as the window length

increases, the WVD converges in probability to its expected value [5] i.e. as  $m \rightarrow \infty$ ,

$$\begin{aligned} W_{2X}(t, \omega) &\xrightarrow{P} E W_{2X}(t, \omega) \\ &\rightarrow \mu_A^2 \delta(\omega - \omega_t) + o(1) \quad (11) \end{aligned}$$

Eq. (11) shows that, for sufficiently large window lengths, the WVD exhibits a peak at the IF  $\omega = \omega_t$ . This result leads to the WVD-based IF estimator,

$$\hat{\omega}_t = \arg \max_{\omega} W_{2X}(t, \omega), \quad t = 0, \dots, n-1 \quad (12)$$

Note that when  $\mu_A = 0$ , the WVD will not exhibit a peak at  $\omega = \omega_t$ .

## 2.2. Cyclic Moments-based Estimator

Shamsunder *et al.* [12] proposed a procedure for estimating the phase parameters of  $s_t$  using estimates of the cyclic moments of  $X_t$ . We provide only a brief outline of the estimation procedure here. The interested reader is referred to [12] for further details.

We define the second-order and first-order cyclic moment estimators, respectively, as

$$\hat{\mathcal{M}}_{2X}(\alpha; \tau) = 1/n \sum_{t=0}^{n-\tau-1} x_t^* x_{t+\tau} \exp(-j\alpha t) \quad (13)$$

$$\hat{\mathcal{M}}_{1X}(\alpha) = 1/n \sum_{t=0}^{n-1} x_t \exp(-j\alpha t) \quad (14)$$

where  $\tau$  is an arbitrary positive integer,  $0 < \tau < n-1$ . Estimates of the phase parameters  $b_1$  and  $b_2$  and the IF  $\omega_t$  can be obtained from the second-order and first-order cyclic moment estimators using the procedure of Table 1.

Table 1: Cyclic Moments-based IF Estimator

1. Estimate  $b_2$  as

$$\hat{b}_2 = 1/(2\tau) \arg \max_{\alpha} |\hat{\mathcal{M}}_{2X}(\alpha; \tau)|^2. \quad (15)$$

2. Form the signal  $x_t^{(1)} = x_t \exp(-j\hat{b}_2 t^2)$  and estimate  $b_1$  as

$$\hat{b}_1 = \arg \max_{\alpha} |\hat{\mathcal{M}}_{1X^{(1)}}(\alpha)|^2. \quad (16)$$

3. Estimate the IF as

$$\hat{\omega}_t = \hat{b}_1 + 2\hat{b}_2 t, \quad t = 0, \dots, n-1. \quad (17)$$

In order to avoid aliasing the phase parameters must satisfy:

$$|b_2| < \pi/(2\tau), \quad |b_1| < \pi \quad (18)$$

## 3. STATISTICAL ANALYSIS

In this section we perform statistical analyses of the WVD-based and cyclic moments-based IF estimators under the following assumptions:

A1  $A_t$  is a stationary real-valued random process with  $p$ th order cumulant  $c_{pA}(\xi_1, \dots, \xi_{p-1})$ . It is assumed that

$$\sum_{\xi_1, \dots, \xi_{p-1}=-\infty}^{\infty} |c_{pA}(\xi_1, \dots, \xi_{p-1})| < \infty, \quad p = 1, 2, \dots$$

A2  $W_t, t = 0, \dots, n-1$  are zero-mean independent and identically distributed (iid) complex-valued Gaussian random variables with finite variance  $\sigma_W^2$ .

The results we present below are asymptotic in the sense that they are valid as the sample length  $n \rightarrow \infty$ .

### 3.1. WVD-based Estimator

Under the assumptions A1 and A2, the variance of the WVD-based IF estimator can be found as

$$\text{var}(\tilde{\omega}_t) = \frac{3}{S_W(2m+1)^3} \left( 2 + \frac{2}{S_A} + \frac{1}{S_W} \right) + O(m^{-7/2}) \quad (19)$$

where  $S_A = \mu_A^2/\sigma_A^2$  and  $S_W = \mu_A^2/\sigma_W^2$ . A brief proof of (19) is given in Appendix 1. The following comments can be made regarding (19):

- The variance of  $\tilde{\omega}_t$  is not affected by correlation in the modulating process  $A_t$ .
- The variance of  $\tilde{\omega}_t$  will increase dramatically at either end of the sample i.e.  $t$  close to zero or  $n$ . This follows from the fact that the window length  $2m+1$  decreases as we move away from the centre of the observation interval.
- When  $\mu_A = 0$ ,  $\text{var}(\tilde{\omega}_t)$  goes to infinity indicating that the WVD-based IF estimator breaks down. This is consistent with the analysis of section 2.1.

### 3.2. Cyclic Moments-based Estimator

The covariance matrix of the cyclic moments-based parameter estimators was derived in [7] under the assumption that  $A_t$  are iid. This analysis was extended to the more general case in which  $A_t$  are correlated in [8]. Let  $\hat{b} = (\hat{b}_1, \hat{b}_2)'$  and  $b = (b_1, b_2)'$ . It was shown in [8] that

$$NE(\hat{b} - b)(\hat{b} - b)'N = \frac{2 + 2/S_A + 1/S_W - \nu(\bar{\tau})}{2\bar{\tau}^2(1 - \bar{\tau})^3} \times \frac{3}{S_W} \begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} + \frac{1}{S_W} \begin{pmatrix} 6 & 0 \\ 0 & 0 \end{pmatrix} + O(n^{-1}) \quad (20)$$

where  $N = \text{diag}(n^{3/2}, n^{5/2})$ ,  $\bar{\tau} = \tau/n$ ,  $0 < \tau < n-1$  and

$$\nu(\bar{\tau}) = \begin{cases} 2(\bar{\tau}^2 - 4\bar{\tau} + 1)/(1 - \bar{\tau})^3, & \bar{\tau} < 1/2 \\ 0, & \bar{\tau} \geq 1/2 \end{cases} \quad (21)$$

Using these results we obtain the variance of the cyclic moment-based IF estimator as

$$\begin{aligned} \text{var}(\hat{\omega}_t) &= \text{var}(\hat{b}_1) + 4\text{cov}(\hat{b}_1, \hat{b}_2)t + 4\text{var}(\hat{b}_2)t^2, \quad (22) \\ &= \frac{3}{S_W n^3} \left\{ \left( \frac{2 + 2/S_A + 1/S_W - \nu(\bar{\tau})}{2\bar{\tau}^2(1 - \bar{\tau})^3} \right) \right. \\ &\quad \times (1 - 4t/n + 4t^2/n^2) + 2 \left. \right\} + O(n^{-7/2}) \quad (23) \end{aligned}$$

Eq. (23) shows that the variance of the cyclic moments-based IF estimator is unaffected by correlation in the modulating process and fails when the mean  $\mu_A$  of the modulating process is zero.

We note that variances of the parameter estimators  $\hat{b}_1$  and  $\hat{b}_2$ , and the IF estimator  $\hat{\omega}_t$  depend on the lag  $\tau$  chosen in the second-order cyclic moment. The dependence of  $\text{var}(\hat{b}_i)$ ,  $i = 1, 2$  on  $\tau$  was noted in [6] where it was shown that the variances of  $\hat{b}_1$  and  $\hat{b}_2$  may be minimised by choosing  $\tau = n/2$  provided that  $2/S_A + 1/S_W < 5.46$ . Comparison of (20) and (23) indicates that  $\text{var}(\hat{\omega}_t)$  has the same dependence on  $\tau$  as the variances of  $\hat{b}_1$  and  $\hat{b}_2$ . It follows that  $\text{var}(\hat{\omega}_t)$  will also be minimised by selecting  $\tau = n/2$ .

## 4. COMPARISON

In this section we compare the variances of the WVD-based and cyclic moments-based IF estimators.

We conduct the comparison between the estimators at the mid-point of the sample i.e.  $t = n/2$ . In this case we have, ignoring lower-order terms,

$$\text{var}(\tilde{\omega}_{n/2}) = \frac{3}{S_W n^3} \left( 2 + \frac{2}{S_A} + \frac{1}{S_W} \right) \quad (24)$$

$$\text{var}(\hat{\omega}_{n/2}) = \frac{6}{S_W n^3} \quad (25)$$

We see from (25) that the cyclic moments-based IF estimator exhibits interesting properties for the case  $t = n/2$ . These are described below:

- $\text{var}(\hat{\omega}_{n/2})$  is independent of the variance  $\sigma_A^2$  of the modulating process i.e. the cyclic moments-based IF estimator at  $t = n/2$  is unaffected by the presence of multiplicative noise. This is in contrast to the WVD-based IF estimator which becomes less accurate as  $\sigma_A^2$  increases.
- $\text{var}(\hat{\omega}_{n/2})$  is independent of the lag  $\tau$ . We see from (18) that the range of allowable values of  $b_2$ , and therefore  $\omega_t$ , can be increased by decreasing  $\tau$ . In general, this is not viable since the accuracy of the cyclic moments-based estimator is poor for small values of  $\tau$ . Such considerations do not apply when estimating the IF at the mid-point of the sample since we can choose small values of  $\tau$  with no loss in accuracy.

It is evident from (24) and (25) that, for  $t = n/2$ , the cyclic moments-based IF estimator has lower variance than the WVD-based IF estimator under all noise conditions. The variances of the estimators will be approximately equal under what may be called high signal-to-noise ratio conditions,  $S_A \gg 1$  and  $S_W \gg 1$ .

The theoretical results given above are now confirmed using simulations. The signal of interest  $s_t$  is a second order PPS with phase parameters  $b_0 = \pi/3$ ,  $b_1 = \pi/32$  and  $b_2 = \pi/2500$ . The sample length is  $n = 512$  and the cyclic moment-based IF estimate is computed using  $\tau = n/4$ . The modulating process  $A_t$  is a sequence of iid Gaussian random variables. We estimate the variances of the cyclic moments-based and WVD-based IF estimators using 5000 realisations of the signal (1). Figure 1 shows the theoretical and estimated variances plotted against  $S_A = -5(1)20$  dB for  $S_W = 0(5)15$  dB. A close correspondence between the theoretical and empirical results is evident. We note that for  $S_W = 5, 10$  and  $15$  dB the cyclic moments-based and WVD-based IF

estimators fail at the same value of  $S_A$ . The cyclic moments-based estimator exhibits a slightly better threshold performance for  $S_W = 0$  dB.

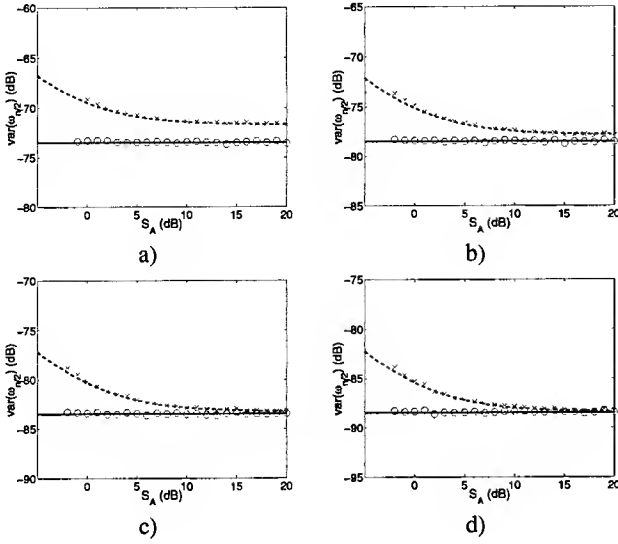


Figure 1: Variances of the WVD-based and cyclic moments-based IF estimators plotted against  $S_A$  for a)  $S_W = 0$  dB, b)  $S_W = 5$  dB, c)  $S_W = 10$  dB, and d)  $S_W = 15$  dB. Theoretical (solid line for cyclic moments and dashed line for WVD) and estimated ('o' for cyclic moments and 'x' for WVD) results are shown.

## 5. DISCUSSION

The problem of estimating the instantaneous frequency (IF) of linear FM signals in real-valued multiplicative and complex white Gaussian additive noise was considered. Theoretical analyses of a cyclic moments-based IF estimator and the Wigner-Ville distribution (WVD)-based IF estimator were performed under the assumptions of a mixing modulating process and complex white Gaussian additive noise. The theoretical results were used to compare the performances of these two estimators at the mid-point of the observation interval. This comparison showed that the cyclic moments-based estimator has a lower variance than the WVD-based estimator for all noise scenarios. The difference between the variances of the two estimators becomes particularly large as the variance of the modulating process increases. This is due to the interesting fact that, at the mid-point of the sample, the variance of the cyclic moments-based IF estimator is independent of the variance of the multiplicative noise.

Future work on this topic will concentrate on extending the analysis to arbitrary order polynomial phase signals.

### A. DERIVATION OF WVD-BASED IF ESTIMATOR VARIANCE

We begin by writing

$$X_{t+\xi} X_{t-\xi}^* = \mu_\xi \exp(j2\omega_t \xi) + Z_\xi, \quad \xi = -m, \dots, m \quad (26)$$

where  $Z_\xi = V_\xi \exp(j2\omega_t \xi)$  and

$$V_\xi = (A_{t+\xi} + U_{t+\xi})(A_{t-\xi} + U_{t-\xi}^*) - \mu_\xi, \quad (27)$$

$$\mu_\xi = \mu_A^2 + c_{2A}(2\xi) + \sigma_W^2 \delta_\xi. \quad (28)$$

Eq. (26) shows that the WVD at time  $t$  can be seen as the discrete Fourier transform of a complex sinusoid with deterministic varying amplitude  $\mu_\xi$  in non-stationary additive noise. Using a generalisation of the result in [10] the IF estimator error can be found as

$$\tilde{\omega}_t - \omega_t = \sum_{\xi=-m}^m \mu_\xi \text{Re}\{d_Z^{(1)}(2\omega_t) - j\xi d_Z^*(2\omega_t)\} / (2\zeta) + O_p(m^{-2}) \quad (29)$$

where

$$\zeta = \sum_{\xi_1=-m}^m \sum_{\xi_2=-m}^m (\xi_1^2 - \xi_1 \xi_2) \mu_{\xi_1} \mu_{\xi_2} \quad (30)$$

$$d_Z(\omega) = \sum_{\xi=-m}^m V_\xi \exp(-j\omega \xi) \quad (31)$$

and  $d_Z^{(v)}(\omega_0)$  denotes the  $v$ th derivative of  $d_Z(\omega)$  with respect to  $\omega$  evaluated at  $\omega = \omega_0$ . Substituting for  $d_Z^{(1)}(2\omega_t)$  and  $d_Z^*(2\omega_t)$  in (29) gives

$$\tilde{\omega}_t - \omega_t = \sum_{\xi=-m}^m \mu_\xi \sum_{\xi_1=-m}^m (\xi_1 - \xi) \text{Im}(V_{\xi_1}) / (2\zeta) + O_p(m^{-2}) \quad (32)$$

Since  $\mathbf{E} V_\xi = 0$ , the estimator  $\tilde{\omega}_t$  is asymptotically unbiased. The variance is

$$\text{var}(\tilde{\omega}_t) = \sum_{\xi_1=-m}^m \sum_{\xi_2=-m}^m \mu_{\xi_1} \mu_{\xi_2} \sum_{\xi_3=-m}^m \sum_{\xi_4=-m}^m (\xi_3 - \xi_1) \times (\xi_4 - \xi_2) \mathbf{E} \text{Im}(V_{\xi_3}) \text{Im}(V_{\xi_4}) / (4\zeta^2) + O(m^{-7/2}) \quad (33)$$

Simple calculations give

$$\mathbf{E} \text{Im}(V_{\xi_3}) \text{Im}(V_{\xi_4}) = \mathbf{E} (V_{\xi_3} V_{\xi_4}^* - V_{\xi_3} V_{\xi_4}) / 2 \quad (34)$$

$$= \{2(\sigma_A^2 + \mu_A^2)\sigma_W^2 + \sigma_W^4\}(\delta_{\xi_3-\xi_4} - \delta_{\xi_3+\xi_4}) / 2 \quad (35)$$

Substituting (35) into (33) gives

$$\text{var}(\tilde{\omega}_t) = \frac{2(\sigma_A^2 + \mu_A^2)\sigma_W^2 + \sigma_W^4}{8\zeta^2} \times \sum_{\xi_1=-m}^m \sum_{\xi_2=-m}^m \mu_{\xi_1} \mu_{\xi_2} \sum_{\xi=-m}^m (2\xi^2 - 2\xi\xi_1) + O(m^{-7/2}) \quad (36)$$

Substituting (28) into (30) gives

$$\begin{aligned} \zeta = & \sum_{\xi_1=-m}^m \sum_{\xi_2=-m}^m (\xi_1^2 - \xi_1 \xi_2) \{\mu_A^4 + \mu_A^2 c_{2A}(2\xi_1) \\ & + \mu_A^2 c_{2A}(2\xi_2) + \mu_A^2 \sigma_W^2 (\delta_{\xi_1} + \delta_{\xi_2}) + c_{2A}(2\xi_1) c_{2A}(2\xi_2) \\ & + c_{2A}(2\xi_1) \sigma_W^2 \delta_{\xi_2} + c_{2A}(2\xi_1) \sigma_W^2 \delta_{\xi_1} + \sigma_W^4 \delta_{\xi_1} \delta_{\xi_2}\} \end{aligned} \quad (37)$$

We will show that only the first term in the summand of (37) is significant. The remaining terms are of lower order and may be ignored for large window lengths. It is straightforward to see that

$$\sum_{\xi_1=-m}^m \sum_{\xi_2=-m}^m (\xi_1^2 - \xi_1 \xi_2) \mu_A^4 = \mu_A^4 (2m+1)^4 / 12 \quad (38)$$

The second term is given by

$$\begin{aligned}
& \sum_{\xi_1=-m}^m \sum_{\xi_2=-m}^m (\xi_1^2 - \xi_1 \xi_2) \mu_A^2 c_{2A}(2\xi_1) \\
&= \mu_A^2 (2m+1) \sum_{\xi=-m}^m \xi \{\xi - (2m+1)/2\} c_{2A}(2\xi) \\
&< \mu_A^2 (2m+1) \sum_{\xi=-m}^m |\xi \{\xi - (2m+1)/2\} c_{2A}(2\xi)| \\
&< \mu_A^2 m(2m+1)^2 \sum_{\xi=-m}^m |c_{2A}(\xi)| \quad (39)
\end{aligned}$$

It follows from the mixing condition of assumption A1 that this second term will be  $O(m^3)$  and so may be ignored compared to the first term which was shown to be  $O(m^4)$  in (38). A similar analysis can be performed for the remaining terms in  $\zeta$  to obtain

$$\zeta = \mu_A^4 (2m+1)^4 / 12 + O(m^3) \quad (40)$$

A similar analysis for the numerator of (36) yields

$$\begin{aligned}
& \sum_{\xi_1=-m}^m \sum_{\xi_2=-m}^m \mu_{\xi_1} \mu_{\xi_2} \sum_{\xi=-m}^m (2\xi^2 - 2\xi\xi_1) = \\
& \mu_A^4 (2m+1)^5 / 6 + O(m^4) \quad (41)
\end{aligned}$$

Substituting (40) and (41) into (36) gives (19).

## 2. REFERENCES

- [1] B. Barkat, B. Boashash, and L.J. Stanković. "Adaptive Window in the PWVD for the IF Estimation of FM Signals in Additive Gaussian Noise". In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP '99*, pages 1317–1320, Phoenix, Arizona, 1999.
- [2] B. Boashash. "Estimating and Interpreting the Instantaneous Frequency of a Signal- Part 1: Fundamentals". *Proceedings of the IEEE*, 80(4):520–538, April 1992.
- [3] B. Boashash and B. Ristic. "Polynomial Time-Frequency Distributions and Time-Varying Higher Order Spectra: Application to the Analysis of Multicomponent FM Signals and to the Treatment of Multiplicative Noise". *Signal Processing*, 67(1):1–23, May 1998.
- [4] M. Ghogho, A.K. Nandi, and A. Swami. "Cramér-Rao Bound and Maximum Likelihood Estimation for Random Amplitude Phase Modulated Signals". *IEEE Transactions on Signal Processing*, 47(11):2905–2916, November 1999.
- [5] E.L. Lehmann. *Theory of Point Estimation*. John Wiley & Sons, New York, 1983.
- [6] M. Morelande. "Optimal Phase Parameter Estimation of Random Amplitude Linear FM Signals using Cyclic Moments". In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP '99*, pages 1557–1560, Phoenix, Arizona, 1999.
- [7] M. Morelande. "Statistical Analysis of Phase Parameter Estimates of Random Amplitude Linear FM Signals". In *Proceedings of the IEEE Signal Processing Workshop on Higher-Order Statistics, HOS '99*, pages 34–38, Caesarea, Israel, 1999.
- [8] M.R. Morelande and A.M. Zoubir. "Performance Analysis of Cyclic Moments-Based Estimators of the Parameters of Random Amplitude Chirp Signals". *IEEE Transactions on Information Theory*, (submitted).
- [9] C.L. Nikias and J.M. Mendel. "Signal Processing with Higher-order Spectra". *IEEE Signal Processing Magazine*, pages 10–37, July 1993.
- [10] B. Porat and B. Friedlander. "Asymptotic Statistical Analysis of the High-Order Ambiguity Function for Parameter Estimation of Polynomial-Phase Signals". *IEEE Transactions on Information Theory*, 42(3):995–1001, 1996.
- [11] P. Rao and F.J. Taylor. "Estimation of Instantaneous Frequency Using the Discrete Wigner Distribution". *Electronics Letters*, 26(4):246–248, February 1990.
- [12] S. Shamsunder, G.B. Giannakis, and B. Friedlander. "Estimating Random Amplitude Polynomial Phase Signals: A Cyclostationary Approach". *IEEE Transactions on Signal Processing*, 43(2):492–505, February 1995.
- [13] G.L. Stuber. *Principles of Mobile Communication*. Kluwer Academic Publishers, 1996.
- [14] H.L. Van Trees. *Detection, Estimation and Modulation Theory. Part 3, Radar-Sonar Signal Processing and Gaussian Signals in Noise*. John Wiley & Sons, 1971.

# THE APPLICATION OF A NONLINEAR INVERSE NOISE CANCELLATION TECHNIQUE TO MARITIME SURVEILLANCE RADAR

Mark R. Cowper and Bernard Mulgrew

Signals and Systems Group,  
Department of Electronics and Electrical Engineering,  
University of Edinburgh, The King's Buildings,  
Edinburgh, EH9 3JL, Scotland, UK.  
mrc@ee.ed.ac.uk, bernie@ee.ed.ac.uk.

## 1. ABSTRACT

A method, which will be referred to as Broomhead's filter method, is reviewed. This method uses a nonlinear inverse to a linear bandstop filter to obtain better noise reduction results, in terms of signal to noise ratio, than linear noise reduction techniques, for the cancellation of wideband chaotic noise from a sinusoid. A novel and unorthodox approach is suggested for the linear bandstop filtering aspect of Broomhead's filter method, which allows it to be applied in situations where the signal of interest has a broader spectrum than that of a sinusoid. This unorthodox approach, referred to as the modified Broomhead filter method, is used to cancel chaotic noise and sea clutter from narrowband Gaussian signals of interest.

## 2. INTRODUCTION

In [1] Broomhead *et al.* proposed using a nonlinear inverse to a linear bandstop filter, in order to cancel wideband chaotic noise from narrowband signals of interest. This technique will be referred to as Broomhead's filter method. In an experiment they carried out, involving a sine wave corrupted by chaotic Ikeda [2] noise, Broomhead *et al.* showed that a radial basis function network (RBFN) [3] nonlinear inverse was able to obtain reasonable performance when the noise process was not known beforehand (*i.e.* the RBFN inverse was trained using the noise corrupted signal of interest). They also described the RBFN inverse as "indispensable", when the noise process was known beforehand (*i.e.* the RBFN inverse was trained using the noise process alone). No linear comparisons were, however, carried out in this work. Strauch [4] carried out the same experiments as those carried out by Broomhead *et al.*, but with linear comparisons. In the experiment with a sine wave corrupted by Ikeda noise, Strauch found that when the noise process was not known beforehand, there was little or no improvement obtained by using a nonlinear inverse, with respect to using a linear inverse. Also, Strauch reported that very little improvement was observed when the noise process was

known beforehand. Strauch reported that both the linear and nonlinear inverses performed more poorly than a 6<sup>th</sup> order Butterworth filter when the noise process was known beforehand, and when it was not. The results for the case when the noise process was known beforehand seemingly contradict the "indispensable" verdict of a nonlinear inverse made by Broomhead *et al.* in such a situation. Due to the seemingly contradictory results of Broomhead *et al.* and Strauch, it was decided to re-investigate the cancellation of Ikeda noise from a sinusoid experiment, using Broomhead's filter method and linear comparisons. The re-investigation of this experiment lead onto a novel modification of Broomhead's filter method, which allowed it to be used for the cancellation of noise from signals of interest which had a broader power spectrum than that of a sinusoid. This modified Broomhead filter method was applied to the cancellation of wideband chaotic Ikeda noise from narrowband Gaussian signals of interest. Finally, the modified filter method was applied to the cancellation of radar sea clutter data from narrowband Gaussian signals. The radar clutter data sets were provided by the Defence Evaluation and Research Agency (DERA) in the UK.

The structure of this paper is as follows. In section 2 a description of the clutter data sets is given. In section 3 a description of Broomhead's filter method is given. In section 4 the cancellation of broadband Ikeda noise from a sinusoid experiment using Broomhead's filter method and linear comparisons is presented. In section 5 the modified Broomhead filter method is discussed, and results are presented for the cancellation of Ikeda noise and sea clutter from narrowband Gaussian signals.

## 3. SEA CLUTTER DATA

### 3.1. Data collection method

A stationary land-based radar was operated in a dwelling mode, that is, with the antenna pointing towards a patch of the sea surface along a fixed direction.

### 3.2. The wavetank sea clutter data sets

The wavetank data sets were recorded in April 1998 as part of an experiment conducted by DERA Malvern and

---

This work was supported by BAE Systems, DERA, and EPSRC.

Racal Radar Defence Systems, at the large wavetank facility, in the ocean engineering laboratory of the University of California, Santa Barbara. The radar used was the Racal-Thorn mobile instrumented data acquisition system (MIDAS). The wavetank is 53m long, 4.26m wide, and 2.13m deep. The wind tunnel extends 30.5m down the tank, leaving an open test section of 22.5m. A wooden beach at the test end of the tank reduces reflections. The wind tunnel can produce wind speeds of up to  $12\text{ms}^{-1}$ . The MIDAS radar used pulse compression. Pulse compression is a signal processing technique which allows a radar to use a long pulse to obtain a large radiated energy, but which also allows the range resolution of a short pulse to be achieved [5]. The range resolution of the radar was 0.3m (*i.e.* an effective pulse width of 2ns). Data was collected in 32 range cells, during wind speeds of  $4\text{ms}^{-1}$  through to  $12\text{ms}^{-1}$ , in steps of  $1\text{ms}^{-1}$ . Pulse to pulse transmit frequency agility was used, in a known (*i.e.* not randomised) sequence. The radar has a dual-polarised receiver. Only the transmit horizontal, receive horizontal (HH) data sets were made available for the work reported in this paper. The effective pulse repetition frequency (PRF) of the radar was 1kHz. The grazing angle and beamwidth were  $6^\circ$  and  $5^\circ$ , respectively. There were 30,000 complex (*i.e.* coherent) samples collected in each range cell, for each wind speed data set.

### 3.3. The Dawber sea clutter data sets

These data sets were collected during experiments conducted by DERA in January and February of 1994, 1995, and 1996, at Sennen Cove near Lands End, and also at Portsmouth (looking at the Isle of Wight) in December 1996. The radar used was the multi-band pulsed radar (MPR) designed and built by Roke Manor. Two data sets from the experiments mentioned above were made available for the work reported in this summary. Both of these data sets were collected without the use of pulse compression or polarisation agility. For both data sets the radar range resolution was 150m (*i.e.* a pulse width of  $1\mu\text{s}$ ), and the PRF was 20kHz. The first data set, which will be called the Dawber-VV data set, was collected using vertical polarisation on transmit and receive during a wind speed of  $12.8\text{ms}^{-1}$ . The second data set, which will be called the Dawber-HH data set, was collected during a wind speed of  $15.4\text{ms}^{-1}$ . The grazing angle and beamwidth used in the collection of both data sets were  $0.12^\circ$  and  $6^\circ$ , respectively. There were 25,600 complex samples collected in each data set: these samples correspond to the temporal signal collected in one range cell, at a distance of 4km from the radar.

## 4. BROOMHEAD'S FILTER METHOD

### 4.1. Diagram and explanation

A diagram of the filtering method proposed by Broomhead *et al.* [1] is given in Figure 1. A block diagram of Broomhead's filtering method is shown in Figure 1(a), with a spectral representation of the filtering operations employed in this filtering technique shown in Figure 1(b). The input signal into the Broomhead filter  $\{b(n)\}$  is a linear combination of a narrowband signal of interest  $\{t(n)\}$  and a wideband noise process  $\{x(n)\}$ , *i.e.*  $b(n) = t(n) + x(n)$ . It is assumed

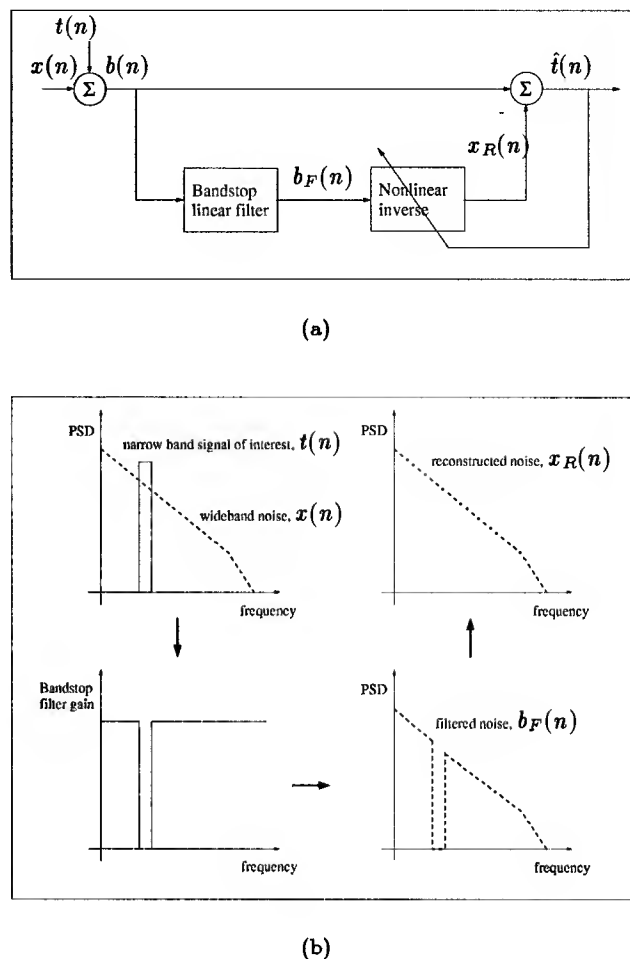


Figure 1: Broomhead's filter method, (a) block diagram, (b) spectral representation of the noise reconstruction process.

that the spectral properties (*i.e.* the band limits) of the signal of interest are known. In a radar context, the noise process could be sea clutter, and the signal of interest could be reflections from a ship or iceberg. A bandstop linear filter is used to remove the signal of interest from the input signal  $b(n)$ . The nonlinear inverse network is used to reconstruct the noise process  $\{x(n)\}$  from the output of the bandstop filter  $\{b_F(n)\}$ . The reconstructed noise process  $\{x_R(n)\}$  at the output of the nonlinear inverse can then be subtracted from the input signal  $\{b(n)\}$  to obtain an estimate of the signal of interest  $\{i(n)\}$ . Figure 1(b) depicts a case where the bandstop filter is orthogonal to (*i.e.* completely removes) the signal of interest, and where the nonlinear inverse manages to reconstruct, perfectly, the noise process, so that the signal of interest may be obtained, exactly, with no errors.

### 4.2. Linear bandstop filter

The filtering method depicted in Figure 1, was originally intended for application to chaotic noise cancellation [1].



In such an application the correct choice of bandstop filter to use is very important. Essentially, the filter must remove the signal of interest (ideally completely), and it must also preserve the dynamics of the noise, so that the nonlinear inverse can properly reconstruct the noise. A discussion on the selection of an appropriate bandstop filter for this application is given in [1, 4], however the key points of this discussion are now briefly summarised:

- Using a short enough linear FIR filter can preserve the dynamics of a chaotic signal.
- If the order of a linear FIR filter is too high, the dynamics of a chaotic signal can be changed.
- The higher the order of a bandstop FIR filter, the greater the attenuation is in the stopband.
- An infinite impulse response (IIR) filter changes the dynamics of a chaotic signal, as it has its own associated dynamics.

Clearly, to preserve the dynamics of a chaotic signal, a low order FIR would be preferred. However, there is a tradeoff between signal suppression, and dynamics distortion: a short enough filter may not change the dynamics of a chaotic process, but it may also not adequately suppresses the signal of interest. There is ambiguity associated with what length of filter constitutes one that is short enough, i.e. of low enough order to not change the dynamics of a chaotic process. There appears to be no clear cut method for selecting an appropriate order, other than to try a simple trial and error approach, to find a suitable filter length which not only adequately suppress the signal of interest, but also does not change the dynamics of the noise process. It should be pointed out that filtering of any kind will distort the dynamics of a chaotic process, however, the aim is to limit this distortion as much as possible so that the nonlinear inverse can produce a reasonable reconstruction of the original chaotic noise process  $\{x(n)\}$  from the bandstop filtered chaotic process  $\{b_F(n)\}$ .

For maritime surveillance radar [6], the *noise*<sup>1</sup> process is sea clutter, which is the term for radar returns from the sea surface. Haykin and Puthusserypady [7] have presented evidence to suggest that sea clutter is a chaotic process. However, evidence presented in [8] has suggested that each clutter data set described in section 2 can be modelled as a non-Gaussian stochastic process, which is the accepted type of model for high resolution and/or low grazing angle sea clutter returns [9]. As will be discussed in section 3.3, the application of Broomhead's filter method to non-Gaussian noise processes, rather than just limiting its application to situations where the noise process is chaotic, is perfectly justifiable. In applying Broomhead's filter technique to cases where the noise process is not chaotic, but is instead a non-Gaussian stochastic process, it might be reasonable to assume that a discussion of the preservation of a signal's dynamics is irrelevant to the choice of a suitable bandstop filter. However, the distortion of a chaotic signal's dynamics resulting from (linear) filtering can be seen as a more general change in the nonlinear properties of the

chaotic signal. Furthermore, it is suggested that this idea of distorting the nonlinear properties of a signal by filtering can be extended to any non-chaotic noise process, which has nonlinear properties that would allow Broomhead's filter method to perform better than a conventional linear approach. In other words, it may be necessary to exercise the same caution in the selection of a suitable bandstop filter for the application of Broomhead's filtering technique when the noise process is described as a non-Gaussian stochastic process, as is required when the noise process is chaotic.

#### 4.3. Why a nonlinear inverse as opposed to a linear inverse?

Linear filtering techniques are unable to relate noise components outside the band of interest, with those inside the band, if the noise process is not available both during training of the linear inverse, and also after training. In a situation where the noise process is only available during training<sup>2</sup>, the best a linear inverse noise suppression approach can achieve is to remove all of the out-of-band noise. It still leaves behind the in-band noise, and therefore performs sub-optimally. The interest in using a nonlinear approach is to try and identify a suitable nonlinear relationship that would allow both the in-band and out-of-band noise components to be suppressed, allowing a nonlinear approach to perform better than a linear one.

As already mentioned in section 3.2, it is justifiable to apply Broomhead's filtering technique to the broad class of non-Gaussian signals. The reason for this is now given. If a process may be described as a Gaussian stochastic process (correlated or uncorrelated), then all its frequency components are independent, and no part of its spectrum is related to another part [10], and it is therefore impossible to relate out-of-band noise to in-band noise, when the noise process is not known beforehand. However, for non-Gaussian stochastic signals it may be possible that a nonlinear approach could relate out of band noise to in-band noise, and could therefore be used to eliminate in-band noise, and achieve better noise suppression than a linear approach.

### 5. CANCELLATION OF BROADBAND CHAOTIC NOISE FROM A SINUSOID

Ikedda map noise was added to a sinusoid so that the signal to noise ratio (SNR) was -2.7dB where,

$$\text{SNR} = 10 \log_{10} \left[ \frac{\sigma_{\text{signal}}^2}{\sigma_{\text{noise}}^2} \right] \quad (1)$$

and  $\sigma_{\text{signal}}^2$  is the variance of the signal of interest, and  $\sigma_{\text{noise}}^2$  is the variance of the noise. As a performance benchmark an 18<sup>th</sup> order bandpass Butterworth filter was used on the noise corrupted sinusoid, and the output SNR achieved was 24.5dB. Broomhead's filter method was applied with

<sup>2</sup>This is the case in maritime surveillance radar, where it is assumed sea clutter data can be collected without any target signal present: for instance, the absence of a target signal could be ensured by visually inspecting an area close to the radar.

<sup>1</sup>Not to be confused with thermal white noise.

and also without the signal of interest present during training of the nonlinear inverse, using a normalised radial basis function network (NRBFN) [11] inverse (with Gaussian kernel functions), and the following bandstop filters: a notch filter [12], an IIR filter, an FIR filter with 25 taps, and an FIR filter with 193 taps. It was found that Broomhead's nonlinear inverse filter method and the linear inverse comparison performed more poorly in terms of output SNR than the Butterworth filter when the FIR and IIR filters were used as the bandstop filter. Furthermore, the nonlinear and linear inverse techniques also performed more poorly than the Butterworth filter when the signal of interest was present during training and the notch filter was used as the bandstop filter. However, the nonlinear inverse outperformed both the linear inverse and the Butterworth filter when the notch filter was used as the bandstop filter and the signal of interest was not present during training (i.e. the noise alone was used to train the inverse), see Figure 2. These results confirm those obtained by Broomhead *et al.*, and contradict those obtained by Strauch.

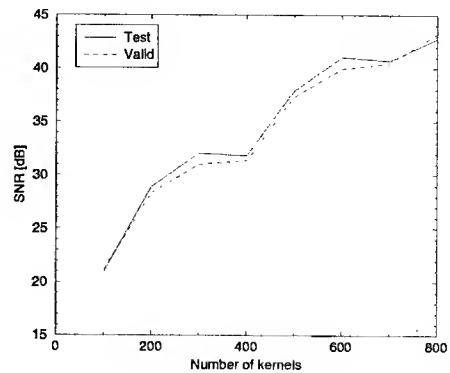
## 6. MODIFIED BROOMHEAD FILTER METHOD

### 6.1. Novel bandstop filtering approach

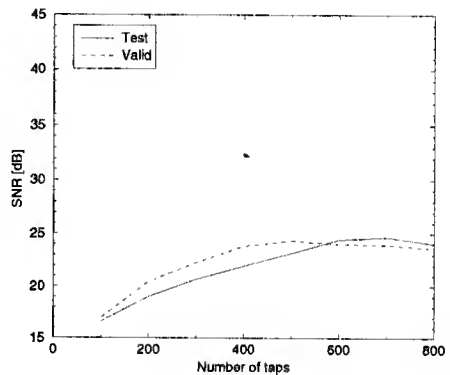
The results discussed in section 4 suggest that the only linear bandstop filtering method which does not distort the dynamics (or nonlinear properties) of the chaotic data too much, and which therefore allows Broomhead's filter method to perform better than linear alternatives, is the notch filter. Therefore, the novel and unorthodox approach for the bandstop filtering aspect of Broomhead's filter method was to use a series of notch filters in order to allow Broomhead's filter method to be applied when the signal of interest has a broader spectrum than that of a sine wave. This modified Broomhead filter approach was applied to the cancellation of Ikeda noise from narrowband Gaussian signals, and to the cancellation of sea clutter from narrowband Gaussian signals.

### 6.2. Cancellation of Ikeda noise from a narrowband Gaussian signal

White Gaussian noise was passed through a 6<sup>th</sup> order IIR Butterworth bandpass filter with a passband from normalised frequency  $f/f_s=0.26$  to  $f/f_s=0.285$ , to produce a narrowband Gaussian signal. Ikeda noise was added to this signal, and the resulting SNR was -8.5dB. As a benchmark performance measure, a 6<sup>th</sup> order Butterworth bandpass filter with a passband of  $f/f_s=0.258$  to  $f/f_s=0.287$  was used, and it achieved an output SNR of -1dB. A NRBFN with an embedding dimension of 4 and an embedding delay of 1 sample was used as the nonlinear inverse in Broomhead's filter. The bandstop linear filter, used to cancel the Gaussian signal, comprised of 3 notch filters (in cascade) with notches at  $f/f_s=0.265$ , 0.275, and 0.285. Nonlinear and linear inverse results are shown in Figure 3. As can be seen from Figure 3, the nonlinear inverse method was found to achieve a better output SNR than both the linear inverse and the bandpass Butterworth filter.



(a)

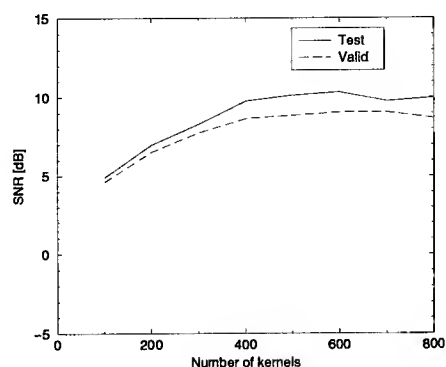


(b)

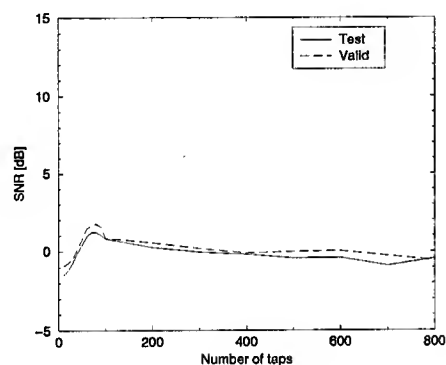
Figure 2: Testing and validation data set output SNR's for (a) nonlinear inverse, (b) linear inverse using a notch filter as the bandstop filter. Inverses trained on noise alone. Training, testing and validation data sets of length 2500 samples were used. An embedding dimension of 4 and an embedding delay of 1 sample were used by the nonlinear inverse.

### 6.3. Cancellation of sea clutter from a narrowband Gaussian signal

White Gaussian noise was passed through a 6<sup>th</sup> order IIR Butterworth bandpass filter with a passband from normalised frequency  $f/f_s=0.0$  to  $f/f_s=0.025$ , to produce a narrowband Gaussian signal. The wavetank  $12ms^{-1}$  gate 14 amplitude data set was added to this signal, and the resulting signal to clutter ratio (SCR) was -2.7dB. A NRBFN was used as the nonlinear inverse in Broomhead's filter method. Embedding dimensions of 4 to 20 in steps of 1 were used. For each embedding dimension, the number of kernels was varied from 100 to 800 in steps of 100. The training length for each simulation was 2500 samples. The bandstop linear filter, used to cancel the Gaussian signal, comprised of 3 notch filters (in cascade) with notches at  $f/f_s=0.005$ , 0.015, and 0.025. A 10 tap linear inverse comparison was



(a)



(b)

Figure 3: Testing and validation data set output SNR's for (a) nonlinear inverse and (b) linear inverse cancellation of Ikeda noise from a narrowband Gaussian signal. Inverses trained on noise alone. Training, testing and validation data sets of length 2500 samples were used.

used with a training length of 2500 samples. The training of the nonlinear and linear inverses was done, for all simulations, using clutter only data (*i.e.* not the clutter corrupted signal of interest). Testing data set results for the linear and nonlinear inverses are given in Figure 4. As can be seen from Figure 4, the simple 10 tap linear inverse performed as well as, or better than, the nonlinear inverses. This was determined to be the case for all the DERA clutter data sets [8].

## 7. SUMMARY

Broomhead's nonlinear inverse filter method was shown to outperform linear alternatives for the cancellation of Ikeda noise from a sine wave when the bandstop filter used was a notch filter, and the inverse was trained on noise only data. A modified Broomhead filter was proposed which allowed Broomhead's filter method to be used to cancel noise from signals with a wider spectrum than that of a sine

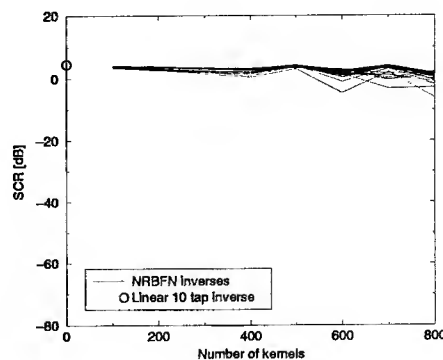


Figure 4: Testing data set output SCR's for nonlinear inverse and linear inverse cancellation of the wavetank  $12ms^{-1}$  gate 14 amplitude data set from a narrowband Gaussian signal. Inverses trained on clutter alone.

wave. The modified Broomhead filter method was shown to achieve better results than linear alternatives in the case of cancellation of chaotic Ikeda noise from a narrowband Gaussian signal, but not in the case of cancellation of DERA sea clutter from a narrowband Gaussian signal.

## REFERENCES

- [1] D. S. Broomhead, J. P. Huke, and M. Potts, "Cancelling deterministic noise by constructing nonlinear inverses to linear filters," *Physica D*, vol. 89, pp. 439–458, 1996.
- [2] H. D. I. Abarbanel, *Analysis of observed chaotic data*. Springer-Verlag, 1996.
- [3] S. Haykin, *Neural Networks: A comprehensive foundation*. Prentice-Hall, 1994.
- [4] P. E. Strauch, *Nonlinear noise cancellation*. PhD thesis, University of Edinburgh, May 1997.
- [5] M. I. Skolnik, *Introduction to radar systems*. McGraw-Hill, second edition ed., 1981.
- [6] K. D. Ward, C. J. Baker, and S. Watts, "Maritime surveillance radar part 1: radar scattering from the ocean surface," *IEE Proceedings Part F*, vol. 137, pp. 51–62, April 1990.
- [7] S. Haykin and S. Puthusserypady, "Chaotic dynamics of sea clutter," *Chaos*, vol. 7, no. 4, pp. 777–802, 1997.
- [8] M. R. Cowper, *Nonlinear processing of non-Gaussian stochastic and chaotic deterministic time series*. PhD thesis, Edinburgh University, Submitted March 2000.
- [9] K. D. Ward and S. Watts, "Radar sea clutter,"  *Microwave Journal*, vol. 28, pp. 109–121, June 1985.
- [10] A. Papoulis, *Probability, random variables, and stochastic processes*. McGraw-Hill, 3rd ed., 1991.
- [11] J. Moody and C. J. Darken, "Fast learning in networks of locally-tuned processing units," *Neural Computation*, vol. 1, no. 2, pp. 281–294, 1989.
- [12] J. G. Proakis and D. G. Manolakis, *Digital Signal Processing*. Macmillan, 1992.

# ADAPTIVE DIGITAL BEAMFORMING RADAR FOR MONOPULSE ANGLE ESTIMATION IN JAMMING

Kai-Bor Yu\* and David J. Murrow\*

\*GE Research & Development Center  
Niskayuna, NY 12309

\*Lockheed Martin Ocean, Radar & Sensors Systems  
Syracuse, NY 13221

## ABSTRACT

This paper describes a digital beamforming architecture for nulling a mainlobe jammer and multiple sidelobe jammers while maintaining the monopulse angle estimation accuracy. It involves two-stage processing using adaptive digital beamforming followed by a mainlobe jammer canceller. A mainlobe jammer blocking matrix and constrained adaptation are employed during the adaptive sidelobe cancellation so that the results of sidelobe jammer cancellation process do not distort subsequent mainlobe cancellation process. This technique is developed to determine the angular location of a target by maintaining the estimation accuracy of the monopulse ratio in the presence of jamming.

## 1. INTRODUCTION

Monopulse is a radar technique in which the angular location of a target can be determined within fractions of a beamwidth by comparisons of two or more simultaneous beams [1]. Monopulse technique for angle estimation fails when there is sidelobe jamming (SLJ) and/or mainlobe jamming (MLJ). If not effectively encountered, electronic jamming prevents successful radar target detection and tracking. We have developed an adaptive beamforming architecture and a signal processing algorithm to cancel mainlobe and sidelobe jammers while maintaining target detection and angle estimation accuracy on mainlobe targets [2]. Our technique makes use of a cascaded scheme where sidelobe jammers are cancelled using adaptive array followed by mainlobe canceller.

In order to motivate this technique, we first review some antenna architectures and adaptive processing schemes for jammer cancellation. Specifically, fully adaptive array [3, 4] and sum-difference mainlobe canceller (MLC)[5] are discussed including their performance in target angle estimation in jamming.

Adaptive array for target detection and angle estimation in

jamming leads to the adaptive sum and difference beams. The sum and difference beams are formed by adaptive receiving array techniques that automatically null the interference sources. Because of the adaptation, the two antenna patterns vary with the external noise field are distorted relative to the conventional sum and difference beams that possess even and odd symmetry, respectively about a prescribed boresight angle. This technique cancels both mainlobe and sidelobe jammers but distort monopulse ratio, thus leading to the bias in angle estimation especially for the situation when jammers are within the mainbeam. This approach for angle estimation works well when the jammers are in the sidelobe.

Applebaum et al [5] have developed a beamforming architecture and algorithm for nulling the MLJ while preserving the monopulse ratio. This technique makes use of the idea that the patterns are separable in azimuth and elevation, i.e. the patterns can be expressed as products of sum and difference factors in azimuth and elevation. We can therefore cancel jammers with nulls along one direction while keeping the non-adapted sum and difference patterns along another direction, thus yielding adapted sum and difference beams with an undistorted monopulse ratio. This technique does not cancel sidelobe jammers.

As the adaptive array works well for sidelobe jammer, and MLC works well for MLJ cancellation, we are motivated to combine these techniques for adaptive monopulse processing. Specifically, we have devised a scheme such that the adaptive array is used for sidelobe cancellation and the MLC is used for mainlobe cancellation. The technique is developed in section 2 and analytical performance evaluation is used to illustrate the technique in section 3.

## 2. COMBINING ADAPTIVE ARRAY AND MAINLOBE CANCELLER

In this section, we present an approach which involves an adaptive digital beamforming (DBF) sub-array followed by an MLC. In order to combine the cancellation technique, we include an appropriate mainlobe maintenance scheme or impose some main-beam constraints in the SLJ canceling process. In this manner, identical nulls are formed at both the sum and difference beams, with the main beams maintained appropriately before applying the MLC.

For the two-stage DBF architecture considered here, there are  $N$  columns in the DBF array, and each column has  $M$  elemental sensors. Partial adaptivity is employed where fixed beamforming is used for each column and adaptive degrees-of-freedom are available along azimuth. In this set up, input for each column is linearly combined to form the column sum and difference beams, i.e.  $\{r_{\Sigma_e}(i), r_{\Delta_e}(i)\}$ ,  $i=1, \dots, N$ . The two sets of beams can then be digitized and linearly sum to form the array sum beam, delta-azimuth beam, delta-elevation beam and delta-delta beam. It should be noted that the quiescent patterns are separable with the following expressions:

$$\begin{aligned} g_{\Sigma}(T_x, T_y) &= g_{\Sigma_a}(T_x) g_{\Sigma_e}(T_y) \\ g_{\Delta_A}(T_x, T_y) &= g_{\Delta_a}(T_x) g_{\Sigma_e}(T_y) \\ g_{\Delta_E}(T_x, T_y) &= g_{\Sigma_a}(T_x) g_{\Delta_e}(T_y) \\ g_{\Delta_{\Delta}}(T_x, T_y) &= g_{\Delta_a}(T_x) g_{\Delta_e}(T_y) \end{aligned} \quad (1)$$

where  $T_x$  and  $T_y$  are the steering directions given by

$$\begin{aligned} T_x &= \cos(EI) \sin(Az) \\ T_y &= \sin(EI) \end{aligned}$$

where  $Az$  and  $EI$  are the azimuth and elevation angle correspondingly. Monopulse ratio along azimuth or elevation direction can then be formed giving azimuth and elevation DOA estimates by using the following:

$$\begin{aligned} f_A(T_x, T_y) &= \frac{g_{\Delta_A}(T_x, T_y)}{g_{\Sigma}(T_x, T_y)} \\ &= \frac{g_{\Delta_a}(T_x)}{g_{\Sigma_a}(T_x)} \end{aligned} \quad (2)$$

$$\begin{aligned} f_E(T_x, T_y) &= \frac{g_{\Delta_E}(T_x, T_y)}{g_{\Sigma}(T_x, T_y)} \\ &= \frac{g_{\Delta_e}(T_y)}{g_{\Sigma_e}(T_y)} \end{aligned} \quad (3)$$

These derivations make use of the separable property of the planar array patterns as given before. In the presence of jamming, the column sum and difference inputs are adaptively weighted as follows:

$$r_{\Sigma} = \hat{w}_{\Sigma}^H r_{\Sigma_e} \quad (4)$$

$$r_{\Delta_A} = \hat{w}_{\Delta_A}^H r_{\Sigma_e} \quad (5)$$

$$r_{\Delta_E} = \hat{w}_{\Delta_E}^H r_{\Delta_e} \quad (6)$$

$$r_{\Delta_{\Delta}} = \hat{w}_{\Delta_{\Delta}}^H r_{\Delta_e} \quad (7)$$

where

$$\hat{w}_{\Sigma\Sigma} = R_{\Sigma_e \Sigma_e}^{-1} w_{\Sigma} \quad (8)$$

$$\hat{w}_{\Delta\Sigma} = R_{\Sigma_e \Delta_e}^{-1} w_{\Delta} \quad (9)$$

$$\hat{w}_{\Sigma\Delta} = R_{\Delta_e \Delta_e}^{-1} w_{\Sigma} \quad (10)$$

$$\hat{w}_{\Delta\Delta} = R_{\Delta_e \Delta_e}^{-1} w_{\Delta} \quad (11)$$

where  $w_{\Sigma}$  and  $w_{\Delta}$  are the nominal sum and difference beamforming weights, and  $r_{\Sigma_e}$  and  $r_{\Delta_e}$  are the column sum and difference beamforming inputs given by:

$$r_{\Sigma_e} = \begin{bmatrix} r_{\Sigma_e}(0) \\ \vdots \\ r_{\Sigma_e}(N-1) \end{bmatrix} \quad (12)$$

$$r_{\Delta_e} = \begin{bmatrix} r_{\Delta_e}(0) \\ \vdots \\ r_{\Delta_e}(N-1) \end{bmatrix} \quad (13)$$

The sample matrix inverses (i.e.  $R_{\Sigma_e \Sigma_e}^{-1}$  and  $R_{\Delta_e \Delta_e}^{-1}$ ) modifies the quiescent weights and corresponds to a nulling preprocessing responsive to jammers. It is essential to include an appropriate mainlobe maintenance technique or to include some constraints in the adaptive process.

After the first stage of adaptive processing, the beams are free of SLJs, but may include the MLJ. The main beams

can then be canceled using an MLC. For example, in order to form the monopulse ratio in elevation, we can adapt the sum and difference beams to cancel the MLJ simultaneously as follows :

$$\hat{r}_{\Sigma_E} = r_{\Sigma} - w_a r_{\Delta_A} \quad (14)$$

$$\hat{r}_{\Delta_E} = r_{\Delta_E} - w_a r_{\Delta_A} \quad (15)$$

This can be done by adapting  $w_a$  in the sum channel and using it in the difference channel or choosing the weight to adapt the sum and difference beams simultaneously by minimizing the sum of output power for both beams. In this way, the monopulse ratio using the adapted sum and difference patterns (i.e.  $\hat{g}_{\Sigma_E}(T_s)$ ,  $\hat{g}_{\Delta_E}(T_s)$ ) can be shown to be preserved along the elevation axis while the jammer is nulled along the azimuth axis as follows:

$$\begin{aligned} \hat{f}_E(T_s) &= \frac{\hat{g}_{\Delta_E}(T_s)}{\hat{g}_{\Sigma_E}(T_s)} \\ &\approx \frac{g_{\Delta_E 0}(T_s) - w_a g_{\Delta_A 0}(T_s)}{g_{\Sigma 0}(T_s) - w_a g_{\Delta_A 0}(T_s)} \\ &= \frac{g_{\Delta_E}(T_s)(g_{\Sigma_A}(T_s) - w_a g_{\Delta_A}(T_s))}{g_{\Sigma_E}(T_s)(g_{\Sigma_A}(T_s) - w_a g_{\Delta_A}(T_s))} \\ &= \frac{g_{\Delta_E}(T_s)}{g_{\Sigma_E}(T_s)} \end{aligned} \quad (16)$$

The same technique can also be used to preserve the monopulse ratio along the azimuth with the mainlobe jammer canceled along the elevation.

### 3. ANALYTICAL PERFORMANCE EVALUATION

In this section, we describe the performance of our technique on monopulse angle estimation in jamming using analytical evaluation. Specifically, we concentrate on the approach of combining the adaptive array and the MLC. Our analytical performance evaluation makes use of a DBF planar array. In this example, the planar array has 28 columns, and each column has 14 elemental sensors, placed half a wavelength apart. Fixed analog beamforming is performed along the elevation for each column, and adaptive digital beamforming capability along the azimuth is available on the resulting column

beams. Uniform nominal weights are used. The null-to-null beam-width is  $8^\circ$  along the azimuth and  $16^\circ$  along the elevation. There are two jammers: one jammer is located within the main-beam ( $2^\circ$  azimuth and  $3^\circ$  elevation), and the other jammer is located at the sidelobe ( $10^\circ$  azimuth and  $2^\circ$  elevation). Both jammers have a jamming-to-noise ratio (JNR) of 45 dB on the element, and a signal-to-noise ratio of 0 dB on the element. We first evaluated the quiescent antenna patterns (Figure 1(a) and (b)). The performance of using adaptive array and MLC with mainlobe maintenance is illustrated in Figure 2. These results show that the adaptive monopulse technique is capable of canceling mainlobe and multiple sidelobe jamming while preserving the radar's ability to estimate the target angle accurately.

### REFERENCES

- [1] D.K. Barton, *Modern Radar System Analysis*, Artech House, 1988.
- [2] K.-B. Yu and D.J. Murrow, *Adaptive Digital Beamforming Architecture and Algorithm for Nulling Mainlobe and Multiple Sidelobe Radar Jammers While Preserving Monopulse Ratio Angle Estimation Accuracy*, U.S. Patent # 5,600,326, Feb. 4, 1997.
- [3] S.P. Applebaum, *Adaptive Arrays*, Syracuse University Research Corp. SPL-769, June 1964.
- [4] R.C. Davis, L.E. Brennan and I.S. Reed, "Angle Estimation with Adaptive Arrays in External Noise Field," *IEEE Trans. on Aerospace and Electronics Systems*, Vol. AES-12, No. 2, March 1976.
- [5] S.P. Applebaum and R. Wasiewicz, *Main Beam Jammer Cancellation for Monopulse Sensors*, Final Tech. Report DTIC RADC-TR-86-267, Dec. 1984.

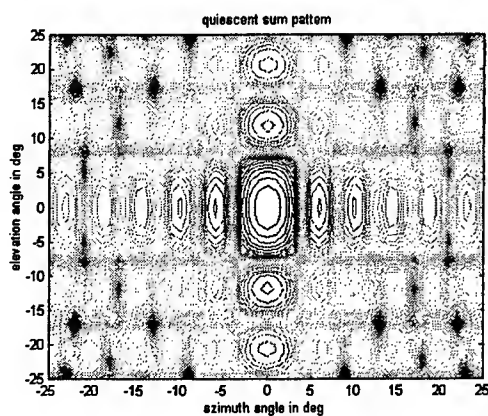


Figure 1(a) Quiescent sum beam pattern

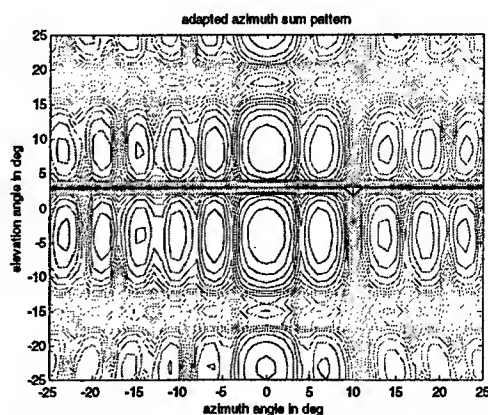


Figure 2(b) Adapted azimuth delta beam pattern

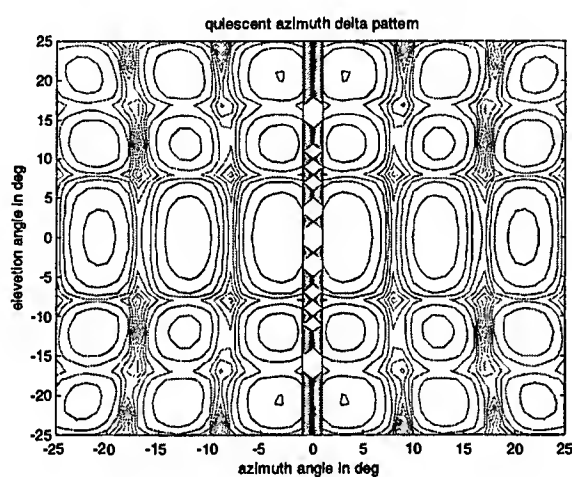


Figure 1(b) Quiescent azimuth delta pattern

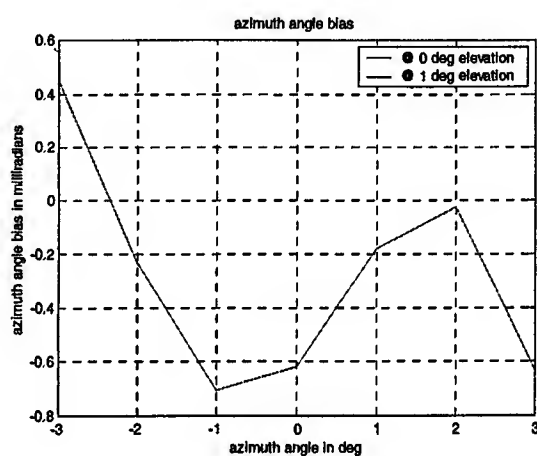


Figure 2(c) Azimuth angle bias

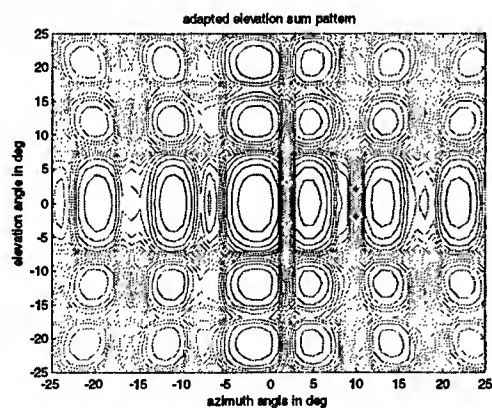


Figure 2(a) Adapted elevation sum beam pattern

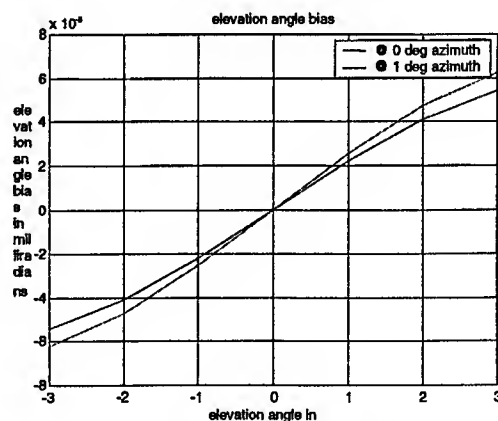


Figure 2(d) Elevation angle bias

# STATISTICAL ANALYSIS OF SMF ALGORITHM FOR POLYNOMIAL PHASE SIGNALS ANALYSIS

*André Ferrari and Gérard Alengrin*

UMR 6525 Astrophysique  
Université de Nice Sophia-Antipolis, Parc Valrose,  
06108 Nice CEDEX 2, FRANCE  
ferrari@unice.fr

## ABSTRACT

The SMF algorithm is designed for the estimation of the coefficients of a constant amplitude polynomial phase signal. It relies on shift invariant signal moments with lower orders than the generalized ambiguity function (GAF) and it does not require maximization. The major contribution of the communication is the derivation of an analytic expression of the SMF error variance for high signal to noise ratios. This result proves the asymptotic efficiency of SMF when a dependency between the number of moments and the number of samples is introduced. Moreover, it underscores the superiority of SMF on GAF with an appropriate choice of the number of moments. Finally, the optimal parameters for order 3 and 4 polynomial phase signal estimation as a function of the signal length are provided.

## 1. INTRODUCTION

This communication is devoted to the estimation of the parameters of a noisy polynomial phase constant amplitude signal. This model is sufficiently general to represent a broad category of real life signals, the reader can refer to [1] for a list of applications relying on this type of signal.

Parametric analysis of polynomial phase signal gave rise to an increasing interest during last years, [2, 7, 4]. The solution generally retained to solve this problem is the Generalized Ambiguity Function (GAF), [5]. The first stage of this method consists in the transformation of the signal in a pure tone. This is achieved iteratively by successive phase differentiations: at each iteration, the differentiation is achieved multiplying the sample at instant  $n$  by the conjugated sample at instant  $n - \tau$ . In the noisy case, the higher degree phase coefficient is estimated from the global maximizer of the transformed signal periodogram. Besides its simplicity, this estimator has the advantage of being asymptotically efficient.

In spite of these advantages, this approach has several limitations. The first is that the signal transformation involves the product of  $2^{M-1}$  signal terms where  $M$  is the degree of the phase. When  $M$  is high, the effect of this large number of terms will be a fast degradation of the algorithm performances. The second is that the method requires a computationally expensive global maximization.

In order to overcome these disadvantages, the SMF algorithm (Stationary Moments Fitting) has been proposed, [3]. The principle of this method relies on the fact that, although the signal is clearly non-stationary, some of its moments are time shift invariant. This property is used to recursively estimate the phase parameters. In [3], the performances of the algorithm are studied using Monte-Carlo simulations. These simulations have shown that when the number of data is small, SMF performances are higher than GAF, [5]. The aim of this communication is to propose an accurate analysis of SMF performances, specifically through a statistical analysis of its precision.

Next section briefly presents the SMF algorithm. An analytic expression of the higher phase coefficient variance is given in the next section. In the third section this result is first validated and then used to establish the asymptotic efficiency of the estimator. Finally, this expression is used to obtain an optimal selection of the algorithm parameters.

## 2. THE SMF ALGORITHM

We assume that  $y_n$  is an order  $M$  noisy polynomial phase signal:

$$y_n = A \exp\{j\phi_n\} + w_n = A \exp\left\{j \sum_{q=0}^M a_q n^q\right\} + w_n, \quad (1)$$



where  $w_n$  is a white and circular iid complex Gaussian noise  $w_n \sim \mathcal{N}_c(0, \sigma_w^2)$  and  $a_0 \sim \mathcal{U}(0, 2\pi)$ .

We have demonstrated in [3] that only moments of order higher or equal to  $2M$  can be time shift invariant. Moreover these statistics only allow the identification of  $a_M$ . Consequently we propose to estimate  $a_M$  from the order  $2M$  stationary moments and similarly as [5] estimate recursively the other parameters. Let  $\mathcal{L}_M$  and  $\mathcal{L}_M^*$  be the two disjoint sets containing the  $M$  delays associated to the unconjugated and conjugated signal samples<sup>1</sup>. The corresponding order  $2M$  "stationary" moment of  $y$  is:

$$\begin{aligned} M_{2M,y}(\mathcal{L}_M, \mathcal{L}_M^*) &\triangleq \mathbb{E} \left\{ \prod_{k=1}^p y_{n+l_k}^{\epsilon_k, n_k} \right\} \\ &= A^{2M} \exp\{(-1)^M j M a_M \prod_{\{k|\epsilon_k=-1\}} l_k^{n_k}\}, \end{aligned}$$

where  $l_1 < l_2 < \dots < l_p$  are the  $p$  different delays,  $n_k$  their multiplicity order and  $\epsilon_k = \pm 1$  indicates if  $y_{n+l_k}$  is conjugated or not. With these notations, we have:

$$\sum_{k=1}^p \epsilon_k n_k = 0. \quad (2)$$

Noticing that  $(r\mathcal{L}_M, r\mathcal{L}_M^*)$ ,  $r = 1, \dots, L$  also lead to stationary moments,  $a_M$  can be estimated by least squares from the angles of these  $L$  moments:

$$\hat{a}_M^{SMF} = \frac{\sum_{r=1}^L r^M \text{angle}(\hat{m}_y(r))}{(-1)^M M \prod_{\{k|\epsilon_k=-1\}} l_k^{n_k} \sum_{r=1}^L r^{2M}}, \quad (3)$$

with:

$$\begin{aligned} \hat{m}_y(r) &\triangleq \hat{M}_{2M,y}(r\mathcal{L}_M, r\mathcal{L}_M^*) \\ &= \frac{1}{N - rl_p} \sum_{k=1}^{N-rl_p} \prod_{q=1}^p y_{k+rl_q}^{\epsilon_q, n_q}. \end{aligned} \quad (4)$$

The set  $(\mathcal{L}_M, \mathcal{L}_M^*)$  will be denoted as the germ of SMF.

### 3. PERFORMANCE ANALYSIS

#### 3.1. Statistical analysis

The next proposition, which is the major contribution of this communication, gives an expression of  $\hat{a}_M$  variance.

<sup>1</sup>Tables containing these delays for various  $M$  are given in [3].

**Proposition 1** For  $SNR \triangleq A^2/\sigma_w^2 \gg 1$ ,

$$\begin{aligned} \text{var}\{\hat{a}_M^{SMF}\} &\approx \frac{1}{2M^2 SNR \prod_{\{k|\epsilon_k=-1\}} l_k^{2n_k}} \\ &\cdot \frac{1}{(\sum_{r=1}^L r^{2M})^2} \left\| \sum_{r=1}^L \frac{r^M}{N - rl_p} \mathbf{u}_r \right\|^2 \end{aligned} \quad (5)$$

where  $\mathbf{u}_r$  is a  $1 \times N$  vector whose non zero components are:

- $\mathbf{u}_r(k) = \sum_{l=1}^q \epsilon_l n_l$  for  $q = 1, \dots, p-1$  and  $1 + rl_q \leq k < rl_{q+1}$ .
- $\mathbf{u}_r(k) = \sum_{l=q+1}^p \epsilon_l n_l$  for  $q = 1, \dots, p-1$  and  $N - r(l_p - l_q) < k \leq N - r(l_p - l_{q+1})$ .

*Proof:* See appendix 1.

- It is worthy to note that the vector  $\mathbf{u}_r$  has only  $2rl_p$  non zero terms. The computation of the norm occurring in the previous expression does not involve a sum of  $N$  terms but of only  $2Ll_p$  terms.
- Consider a cubic phase signal;  $M = 3$ . A solution is the choice of the order 6 germ:

$$\begin{aligned} M_{6,y}(\{0, 3, 3\}, \{1, 1, 4\}) &= \\ &\mathbb{E} \{ y_n y_{n-3}^2 (y_{n-1}^2 y_{n-4})^* \}. \end{aligned} \quad (6)$$

Consequently  $p = 4$  and:

$$\begin{array}{c|c|c|c} l_1 = 0 & l_2 = 1 & l_3 = 3 & l_4 = 4 \\ n_1 = 1 & n_2 = 2 & n_3 = 2 & n_4 = 1 \\ \epsilon_1 = 1 & \epsilon_2 = -1 & \epsilon_3 = 1 & \epsilon_4 = -1 \end{array}$$

The non zero components of  $\mathbf{u}_r$  are:

$$\begin{array}{c|c} 1 \leq k < r+1 & \mathbf{u}_r(k) = 1 \\ r+1 \leq k < 3r+1 & \mathbf{u}_r(k) = -1 \\ 3r+1 \leq k < 4r+1 & \mathbf{u}_r(k) = 1 \\ N-4r < k \leq N-3r & \mathbf{u}_r(k) = -1 \\ N-3r < k \leq N-r & \mathbf{u}_r(k) = 1 \\ N-r < k \leq N & \mathbf{u}_r(k) = -1 \end{array}$$

### 4. SIMULATIONS

#### 4.1. Comparison with Monte-Carlo simulations

The aim of this first simulation is to compare the theoretical expression of the variance given in the previous section with the variance estimated by Monte-Carlo simulations. The simulations have been realised with a cubic phase signal and the germ  $(\{0, 3, 3\}, \{1, 1, 4\})$ .

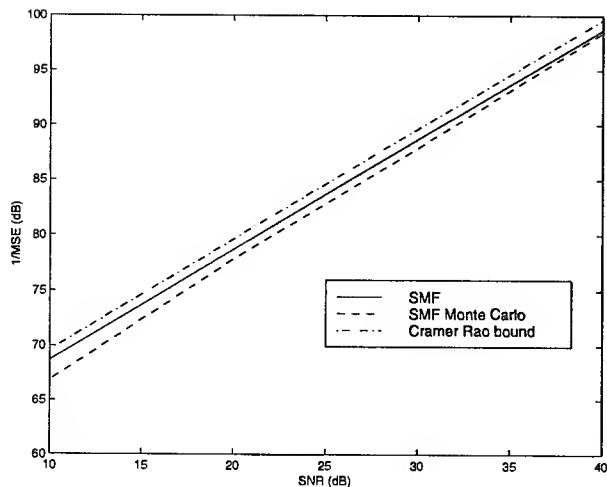


Figure 1: Comparison between theoretical variance and Monte-Carlo simulations.  $M = 3$ , the germ equals  $(\{0, 3, 3\}, \{1, 1, 4\})$ ,  $N = 20$  and  $L = 3$ .

The number of samples is  $N = 20$  and  $L = 3$ . The variances have been estimated from 500 independent realisation of the noise for each signal to noise ratio. Figure 1 shows the good adequation between the estimated variances and the expression (5). The Cramer Rao lower bound (CRLB) of  $\hat{a}_3$  given by [6] is also given in the plot.

#### 4.2. Asymptotic performances

For a given value of  $L$ , (5) shows that the variance of  $\hat{a}_M$  is  $O(1/N^2)$ . The CRLB of  $\hat{a}_M$  is  $O(1/N^{2M+1})$ . Consequently SMF is not an asymptotically efficient estimator for  $L$  constant. However, the expression of  $\hat{m}_y(r)$  given above shows that  $L$  can take values up to  $(N-1)/l_p$ . A possibility to increase the efficiency of SMF is to chose  $L$  as an increasing function of  $N$ .

For example  $L$  can be chosen as  $(N-1)/l_p - \tau$ . In this case  $\tau l_p$  is the minimum number of terms averaged for the estimation of the moments. This approach is analog to the one used in classical spectral analysis where the correlogram is windowed to reduce its variance, [8]. The choice of  $\tau$  relies on a compromise between the reduction of the estimator variance and the poor quality of the last moments caused by a low average of terms in (4). It is important to remember that the moments (4) are always unbiased.

In order to study the performances of the estimator, the ratio

$$\eta_N \triangleq \sqrt{\frac{\text{var}\{\hat{a}_M^{SMF}\}}{\text{CRLB}\{\hat{a}_M\}}} \quad (7)$$

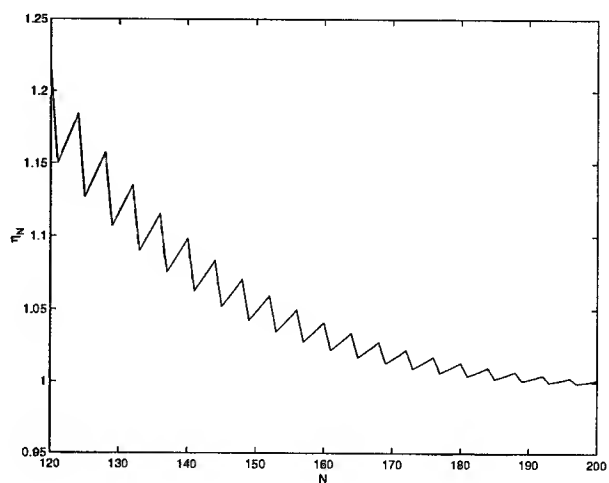


Figure 2: Asymptotic efficiency of SMF.  $M = 3$  and the germ equals  $(\{0, 4, 5\}, \{1, 2, 6\})$ .

is defined for  $SNR \gg 1$ . Note that this ratio is not function of the signal to noise ratio. Figure 2 represents  $\eta_N$  as a function of  $N$  for  $\tau = 11$  and the germ  $(\{0, 4, 5\}, \{1, 2, 6\})$ . This result clearly proves that in this case SMF is asymptotically efficient.

#### 4.3. Comparison with the GAF

In order to compare the performances of SMF and GAF, we define the ratio:

$$\nu_N \triangleq \sqrt{\frac{\text{var}\{\hat{a}_M^{SMF}\}}{\text{var}\{\hat{a}_M^{GAF}\}}} \quad (8)$$

The two expressions of the variance are obtained in the case  $SNR \gg 1$ . The analytical expression of  $\hat{a}_M^{GAF}$  variance has been first computed in [5] and after, with less restrictive conditions, in [1].

Figure 3 represents  $\nu_N$  as a function of  $N$  in the case  $M = 3$  for the germ  $(\{0, 4, 5\}, \{1, 2, 6\})$  and for various values of  $L$ . These plots show that for a given value of  $L$ ,  $\nu_N$  has a minimum smaller than 1. This result allows to select, for a given number of measured samples, an "optimal" value of  $L$ . Moreover, it shows that it is always possible to choose a  $L$  in order that the variance of SMF is lower than the variance of GAF.

#### 4.4. Optimal choice of SMF parameters

This last section provides the optimal values for the SMF parameters in the case of order 3 and 4 polynomial phase signal. Using formula (5), the optimal values of the germ with the associated value of  $L$  have been computed for different values of  $N$ . The results are

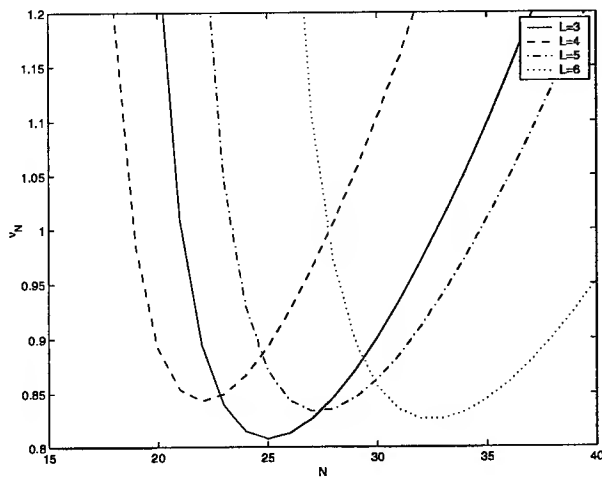


Figure 3: Comparison between SMF and GAF for  $\hat{a}_3$ . The germ equals  $(\{0, 4, 5\}, \{1, 2, 6\})$ .

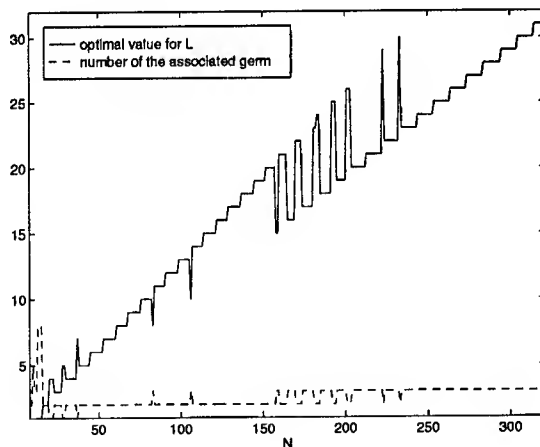


Figure 4: Optimal choice of  $L$  and the germ in the case  $M = 3$ .

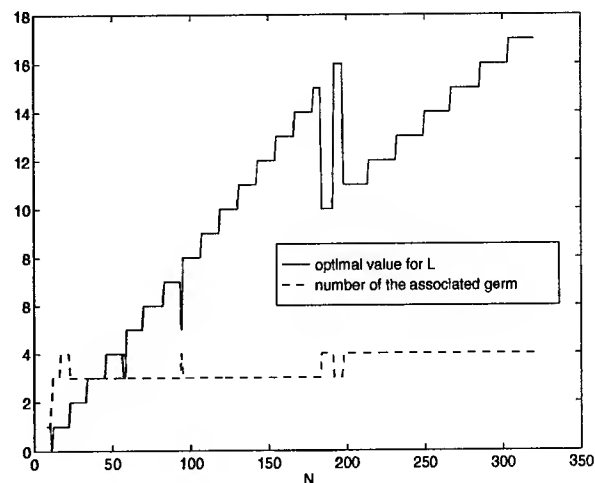


Figure 5: Optimal choice of  $L$  and the germ in the case  $M = 4$ .

provided in figure 4 and 5 which gives for  $N$  in the range  $8 \leq N \leq 314$  the optimal  $L$  and the number of the corresponding germ. The germs are numbered:

#1	$(\{0, 3, 3\}, \{1, 1, 4\})$ ,
#2	$(\{0, 4, 5\}, \{1, 2, 6\})$ ,
#3	$(\{0, 5, 7\}, \{1, 3, 8\})$ ,
#4	$(\{0, 5, 8\}, \{2, 2, 9\})$ ,
#5	$(\{0, 6, 9\}, \{1, 4, 10\})$ ,

for  $M = 3$ , and:

#1	$(\{0, 3, 4, 7\}, \{1, 1, 6, 6\})$ ,
#2	$(\{0, 4, 7, 11\}, \{1, 2, 9, 10\})$ ,
#3	$(\{0, 5, 5, 10\}, \{1, 2, 8, 9\})$ ,
#4	$(\{0, 5, 10, 15\}, \{1, 3, 12, 14\})$ ,
#5	$(\{0, 5, 12, 17\}, \{2, 2, 15, 15\})$ ,

for  $M = 4$ .

This result shows that for the cubic phase signal the optimal germ in the range  $10 \leq N \leq 150$  is #2 =  $(\{0, 4, 5\}, \{1, 2, 6\})$ . This result is in accordance with the one obtained by simulation in [3]. After a short transition interval, the optimal germ becomes #3 =  $(\{0, 5, 7\}, \{1, 3, 8\})$  for the following interval which occurs for  $N > 200$ . The behaviour in the case  $M = 4$  is similar, the optimal germ for  $10 \leq N \leq 200$  being #3 =  $(\{0, 5, 5, 10\}, \{1, 2, 8, 9\})$ . It is worthy to note that in both cases,  $L$  is approximately a linear function of  $N$  in the successive intervals.

## 5. CONCLUSIONS

This communication gives a theoretical expression of the error variance of SMF algorithm for polynomial

phase higher order degree coefficient estimation. This expression leads to a modification of the algorithm in order to obtain its asymptotic efficiency. Moreover it confirms the superiority of the SMF with respect to the GAF, conditioned to a correct choice of the algorithm parameters. Finally, optimal values of the SMF parameters are provided for order three and four polynomial phase signals.

### A. PROOF OF PROPOSITION 1

Denote by  $w = w_r + jw_i$  the noise vector. For a high SNR:

$$\begin{aligned}\hat{a}_M &= G(w_r, w_i) \\ &\approx G(0, 0) + \left. \frac{dG(w_r, w_i)}{dw_r} \right|_{w=0}^t w_r + \left. \frac{dG(w_r, w_i)}{dw_i} \right|_{w=0}^t w_i\end{aligned}$$

If  $w = 0$ , it can be easily verified that  $\hat{m}_y(r) = m_x(r)$  and consequently  $G(0, 0) = a_M$ . Using the circularity of  $w$ , we obtain:

$$\text{var}\{\hat{a}_M\} = \frac{\sigma_w^2}{2} \left( \left\| \left. \frac{dG(w)}{dw_r} \right|_{w=0} \right\|^2 + \left\| \left. \frac{dG(w)}{dw_i} \right|_{w=0} \right\|^2 \right) \quad (9)$$

Herein, the computation of the gradient of  $G(w)$  leads to the computation of the gradient of the angle of  $\hat{m}_y(r)$ . The term associated to the gradient with respect to the real component equals:

$$\begin{aligned}\left\| \left. \frac{dG(w)}{dw_r} \right|_{w=0} \right\|^2 &= \frac{1}{M^2 \prod_{\{k|\epsilon_k=-1\}} l_k^{2n_k} (\sum_{r=1}^L r^{2M})^2} \\ &\cdot \left\| \sum_{r=1}^L r^M \text{Imag} \left( \frac{1}{\hat{m}_y(r)} \frac{d\hat{m}_y(r)}{dw_r} \right) \right\|_{w=0}^2\end{aligned}$$

Substituting (4) in the expression of the  $n$ th component of the previous gradient, we obtain:

$$\begin{aligned}\text{Imag} \left( \frac{1}{\hat{m}_y(r)} \frac{d\hat{m}_y(r)}{dw_{r,n}} \right) \Big|_{w=0} &= \\ \frac{1}{N - rl_p} \sum_{k=1}^{N-rl_p} \text{Imag} \left( \frac{1}{\hat{m}_y(r)} \frac{d \prod_{q=1}^p y_{k+rl_q}^{\epsilon_q, n_q}}{dw_{r,n}} \right) \Big|_{w=0}\end{aligned} \quad (10)$$

The previous derivatives do not equal zero if  $n = k + rl_q$  and in this case :

$$\begin{aligned}\text{Imag} \left( \frac{1}{\hat{m}_y(r)} \frac{d \prod_{q=1}^p y_{k+rl_q}^{\epsilon_q, n_q}}{dw_{r,n}} \right) \Big|_{w=0} &= \text{Imag} \left( \frac{n_q}{x_n^{\epsilon_q}} \right) \\ &= -\frac{1}{A} n_q \epsilon_q \sin \phi_n.\end{aligned}$$

Using (2) and the previous expression, it is possible to compute the expression of the sum in (10) for the different values of  $n$ :

$$\text{Imag} \left( \frac{1}{\hat{m}_y(r)} \frac{d\hat{m}_y(r)}{dw_r} \right) \Big|_{w=0} = -\frac{1}{A(N - rl_p)} \mathbf{S} \mathbf{u}_r, \quad (11)$$

where  $\mathbf{S}$  is a diagonal  $N \times N$  matrix with diagonal terms  $\{\sin \phi_1, \dots, \sin \phi_N\}$ .

A similar development leads to:

$$\text{Imag} \left( \frac{1}{\hat{m}_y(r)} \frac{d\hat{m}_y(r)}{dw_i} \right) \Big|_{w=0} = \frac{1}{A(N - rl_p)} \mathbf{C} \mathbf{u}_r, \quad (12)$$

where  $\mathbf{C}$  is a diagonal  $N \times N$  matrix with diagonal terms  $\{\cos \phi_1, \dots, \cos \phi_N\}$ . The substitution of (11,12) in (9) terminates the proof. ■

### REFERENCES

- [1] M. Benidir and A. Ouldali. Polynomial phase signals analysis based on the polynomial derivatives decompositions. *IEEE Transactions on Signal Processing*, 1998. à paraître.
- [2] P.M. Djurić and S.M. Kay. Parameter Estimation of Chirp Signals. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 38:2118–2126, 1990.
- [3] A. Ferrari, C. Theys, and G. Alengrin. Polynomial phase signal analysis using stationary moments. *EURASIP Signal Processing*, 54:239–248, nov 1996.
- [4] R.M. Liang and K.S. Arun. Parameter Estimation for Superimposed Chirp Signals. In *IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 273–276, 1992.
- [5] S. Peleg and B. Friedlander. The discrete polynomial-phase transform. *IEEE Transactions on Signal Processing*, 43(8):1901–1912, August 1995.
- [6] S. Peleg and B. Porat. The Cramer-Rao Lower Bound for signals with constant amplitude and polynomial phase. *IEEE Transactions on Signal Processing*, 39(3):749–752, March 1991.
- [7] S. Peleg and B. Porat. Estimation and Classification of Polynomial-Phase Signals. *IEEE Transactions on Information Theory*, 37(2):1991–430, mar 1991.
- [8] Petre Stoica and Randolph Moses. *Introduction to Spectral Analysis*. Prentice Hall, 1997.

# PASSIVE SONAR SIGNATURE ESTIMATION USING BISPECTRAL TECHNIQUES

R.K. Lennartsson,<sup>1</sup> J.W.C. Robinson,<sup>1</sup> L. Persson,<sup>1</sup> M.J. Hinich<sup>2</sup> and S. McLaughlin<sup>3</sup>

<sup>1</sup>Defence Research Establishment, SE 172 90 Stockholm, Sweden.

e-mail: {ron,john,leifp}@sto.foa.se

<sup>2</sup>Applied Research Laboratories, University of Texas at Austin, Austin, TX 78713-8029.

e-mail: hinich@mail.la.utexas.edu

<sup>3</sup>Dept. of Electronics and Electrical Eng. University of Edinburgh, Edinburgh EH9 3JL, UK.

e-mail: sml@ee.ed.ac.uk

## ABSTRACT

An important task in underwater passive sonar signal processing is the determination of target signatures based on the narrow-band signal content in the received signal. To achieve good classification performance it is important to be able to separate the different sources (e.g. engine, hull and drive) present in the signature, and to determine the distinct frequency coupling pattern of each of these sources. In this work we demonstrate how this can be done using bispectral techniques applied to data recorded at a sea trial in the Baltic Sea. As a target we used a 23 ft fiberglass motor boat powered by a 4-cylinder, 4-stroke, turbo-charged diesel engine connected to a stern drive with two counter rotating propellers. Data was recorded with a bottom mounted hydrophone array as well as with accelerometers mounted on the engine and hull. It was found that the harmonics that propagated through water are engine related at low speeds and drive related at high speeds. The hull vibrations are only present at very low speeds. Moreover, we found that normalized bispectrum measures (skewness) could provide additional coupling information not visible in the standard bispectrum.

## 1. INTRODUCTION

In passive sonar signature estimation it is important to be able to separate the different narrow-band contributions that are present in the received signal. If the source is a vessel with a conventional engine/drive configuration a good first characterization can often be obtained with the power spectrum alone, but for a more precise characterization the *phase couplings* between harmonics must be uncovered. The phase coupling patterns can be used to *separate the different sources*

present in the signature. However it is a well-known fact that conventional power spectral techniques are phase-blind and cannot be used to track phase couplings, hence the use of *bispectral techniques* [1].

A stationary signal with narrow-band content at the frequencies  $f_1, f_2$  and  $f_1 + f_2$  will show peaks in the power spectrum at these frequencies. In the bispectrum however, a peak at the bifrequency  $(f_1, f_2)$  will occur if and only if the signals are phase-coupled. The ability of the bispectrum to detect phase-couplings has been utilized in such diverse areas as diagnosis of heart conditions [2], nonlinear wave interaction in tidal waves [3], and machine monitoring [4].

In the present work we report on an experiment on harmonic characterization of (hydro-)acoustic signals performed in shallow waters using a small motor boat with a diesel engine and a stern drive as a source. Our main objective has been to determine if the phase coupling pattern between harmonics present in the engine, drive and hull are preserved after propagation through (shallow) water, and if it is possible to utilize this information for classification purposes. More specifically, the focus has been on the possibility to separate the different generating sources (engine, drive and hull) in hydrophone data. A secondary objective has been to compare bispectrum and skewness based techniques in this particular application.

## 2. SPECTRUM, BISPECTRUM AND SKEWNESS

Given a discrete time series  $x(n)$  obtained by sampling with frequency  $f_s$  a (zero-mean, second-order stationary) process  $x(t)$ . The power spectrum  $P(k)$  of  $x(n)$ , for discrete frequency  $f = k\Delta f$  with  $\Delta f = f_s/M$ , can be estimated by conventional averaging of peri-

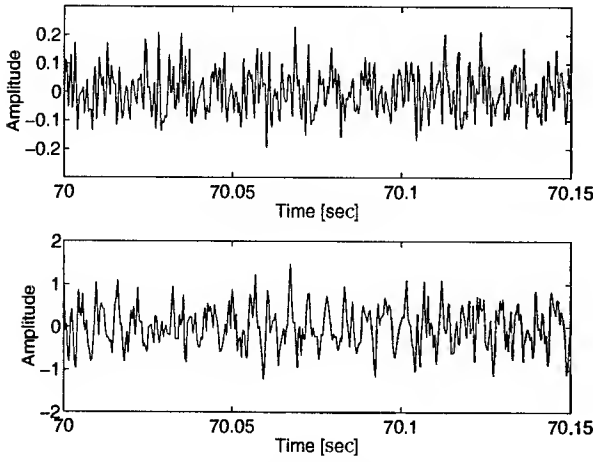


Figure 1: Typical time series of the signals; engine accelerometer (top) and hydrophone (bottom).

odograms by dividing the time series  $x(n)$  into  $K$  (possibly overlapping) blocks, each of length  $M$ , and computing (possibly using tapering) the  $M$  point DFT for each block. The estimated power spectrum  $\hat{P}(k)$  at frequency bin  $k$  is then obtained as

$$\hat{P}(k) = \frac{1}{K} \sum_{j=1}^K |X_j(k)|^2,$$

where  $X_j(k)$  is the DFT over block  $j$  at frequency bin  $k$ . In a similar way (assuming third-order stationarity) the bispectrum  $B(k, \ell)$  and skewness<sup>1</sup>  $S^2(k, \ell)$ , for discrete bifrequency  $(f_1, f_2) = (k\Delta f, \ell\Delta f)$ , can be estimated with the direct method. The (averaged) bispectrum and skewness estimates  $\hat{B}(k, \ell)$  and  $\hat{S}^2(k, \ell)$ , respectively, are then given by [5]

$$\hat{B}(k, \ell) = \frac{1}{K} \sum_{j=1}^K X_j(k) X_j(\ell) X_j^*(k + \ell),$$

$$\hat{S}^2(k, \ell) = \frac{|\hat{B}(k, \ell)|^2}{\hat{P}(k) \hat{P}(\ell) \hat{P}(k + \ell)},$$

where  $(\cdot)^*$  denotes complex conjugation.

### 3. SEA TRIAL

The sea trial was conducted in the Baltic Sea off the east coast of Sweden, in shallow waters of approximately constant depth, 30 meters. As target a 23 ft fiberglass motor boat (Botnia Marine model Targa-23) was used, powered by a 4-cylinder, 4-stroke turbocharged Volvo Penta (VP) diesel engine (type AD31P-A) equipped with a VP Aquamatic stern drive (type

<sup>1</sup>A square-root of it is called bicoherence index in [1].

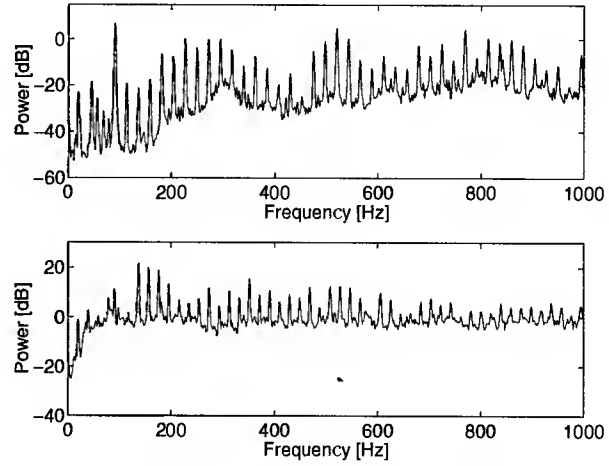


Figure 2: Typical power spectra of the signals; engine accelerometer (top) and hydrophone (bottom).

AD31/DP) having an engine/drive gear ratio of 2.3:1. The drive was fitted with two counter-rotating propellers (VP type A7) having 3 (front) and 4 (rear) blades, respectively.<sup>2</sup> Engine and hull vibrations were recorded with two one-axis accelerometers, one fitted directly on the engine mount and one to the hull close to the engine. Water-propagated sound was recorded using a hydrophone array, with four wide-band omnidirectional hydrophones horizontally equally spaced, but at various depths. In the subsequent analysis presented in this work one hydrophone mounted at a depth of 17 meters was utilized.

In order to separate the different sources involved (e.g. engine, hull and drive), several recordings were made at various rpm, both with the boat drifting freely with the drive disconnected and with the boat moving with the drive connected. In each of the recordings where the boat was powered by its drive it was run on a straight track, at constant throttle, passing directly above the hydrophone array. Ambient sea noise was recorded and analyzed to ensure that it had negligible effect on the end result. The weather conditions during the sea trial were good with wind speeds below 5 m/s. The sound velocity profile was also measured, and was found to be approximately flat over the whole water depth. All data was recorded with a sampling rate of 25 kHz, which was considered to be sufficiently high since most signal and noise power was below 5 kHz and virtually no power was present over 10kHz. To ensure that the phase relations in the recorded signal

<sup>2</sup>The most notable advantage with having two counter-rotating propellers, rather than one single, is less noise and vibration. Hence, the power in drive related signals from this vessel can be expected to be lower than with other forms of drives.

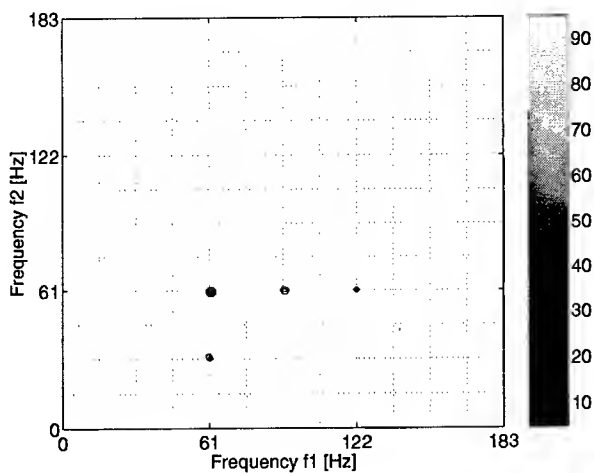


Figure 3: Bispectrum magnitude for the hull accelerometer time series from the 1830 rpm straight-track run (at CPA). The grid spacing is 15.25 Hz which corresponds to half the engine axis rotation frequency.

and noise were preserved no prefiltering was used.

#### 4. DATA ANALYSIS AND RESULTS

In the following we will show the results of an analysis of time series from three different rpm straight-track recordings, at 1830, 2712 and 3549 rpm, respectively. The data and results presented here are taken from a time frame of 15 seconds duration around the closest point of approach (CPA) to the hydrophone array. In the bispectrum estimates the number of blocks was  $K = 22$  and the number of points in the DFTs was  $M = 16384$ . The same number of points in the DFTs were used in the skewness estimates but to achieve a consistent estimate an overlap of 12288 was used yielding a total number of blocks  $K = 88$ . A Hamming tapering was applied to data in all DFT computations.

Figure 1 displays a typical example of time series from the hull accelerometer and the hydrophone. The corresponding power spectra of the time series in Fig. 1 are seen in Fig. 2. By conventional power spectral based analysis it is difficult to separate and relate the peaks of different sources (engine, drive and hull). However, with bispectral analysis it is easier to identify and separate the sources.

##### 4.1. Bispectrum

In Figure 3 the estimated absolute value of the bispectrum for the hull accelerometer data from the 1830 rpm straight-track run at CPA is displayed. The grid spacing is 15.25 Hz which corresponds to half the engine

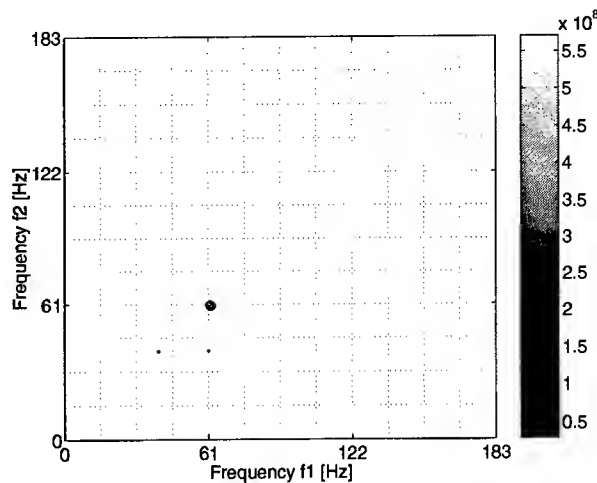


Figure 4: Bispectrum magnitude for the hydrophone time series from the 1830 rpm straight-track run (at CPA). The grid spacing is the same as in Fig. 3.

axis frequency. Here, and in all subsequent figures involving the bispectrum, the values are quantized to 10 levels with white indicating the lowest level and the other levels given by the grayscale on the right. It can be seen that there are strong coupled modes, induced mainly by the torque variations of the engine due to inertia, piston angular velocity and gas pressure variations. This is an example of the "second order" harmonics appearing at twice the engine axis rotation frequency [6], which are moreover coupled to the associated fourth order harmonics. Coupled second and third order engine axis harmonics are also visible. The corresponding bispectrum for the hydrophone is seen in Fig. 4 (same grid spacing as in Fig. 3) where the only visible (off-diagonal) peak is at bifrequency (approx) (61, 40) Hz, which represents a coupling between an engine and a drive harmonic.

In Figures 5 and 6 the bispectra for the hull accelerometer and hydrophone data, respectively, for the 2712 rpm straight-track run is shown. The grid used in Fig. 5 is 22.6 Hz, which is half the engine axis frequency. Also one can see strong coupled engine harmonics in the hull accelerometer data, at the second and third order. Moreover, one can see couplings between the engine and drive, at bifrequencies (approx) (117, 20) Hz and (136, 20) Hz. The grid spacing used in Fig. 6 is 19.5 Hz, which corresponds to the propeller axis frequency, and several frequency couplings are visible. Notable in particular is the peak at bifrequency (approx) (137, 20) Hz (and its neighbors), and the band-like structure of peaks around  $f_2 = 156$  Hz. It appears that all the peaks fall on the grid. Hence, these harmonics are drive related. This can be explained by

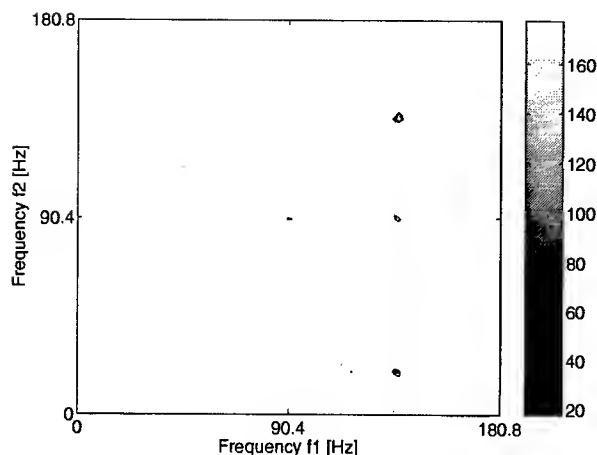


Figure 5: Bispectrum magnitude for the hull accelerometer time series from the 2712 rpm straight-track run (at CPA). The grid spacing is 22.6 Hz, which is half the engine axis frequency.

the fact that the propeller noise increases dramatically when the speed increases and that the engine load, and hence vibrations, are lower at the higher speed since the boat is then hydrofoiling.

In Figures 7 and 8 the bispectra of the hull accelerometer and hydrophone data, respectively, from the 3549 rpm straight-track run are displayed. The results are about the same as the ones obtained at 2712 rpm. In the accelerometer bispectrum in Fig. 7, where the grid spacing is 29.6 Hz corresponding to half the engine axis frequency, again one can see a few engine-engine coupled harmonics, at the expected orders, and some additional engine-drive coupled harmonics. In the hydrophone bispectrum in Fig. 8, where the grid spacing is 25.6 Hz, which corresponds to the propeller axis frequency, a very rich coupling structure is again visible. Also here it appears as if all peaks fall on the grid and hence the harmonics are all drive related.

#### 4.2. Skewness

Given the fact that apparently all visible coupled harmonics in the hydrophone data for higher speeds (2712 and 3549 rpm) fall on frequencies commensurable with the drive frequency a natural question is if a more careful analysis, using for instance the skewness, would reveal additional coupling information. This indeed turns out to be the case, as shown in Fig. 9 where the skewness for the 2712 rpm straight-track run is shown using a grid spacing of 19.5 Hz, which corresponds to the propeller axis frequency. Here, only the values exceeding half of the full range are shown, and these values

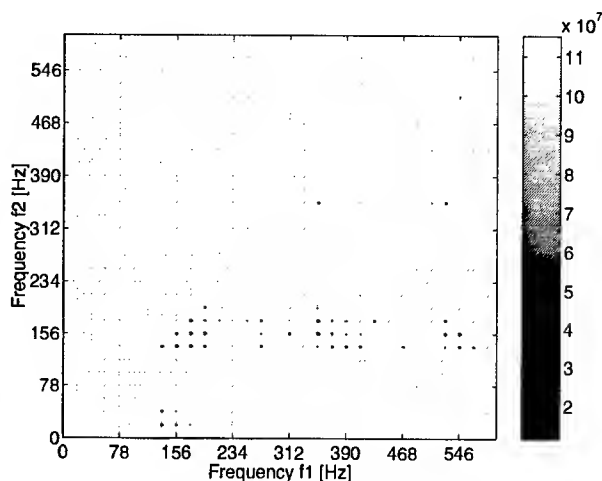


Figure 6: Bispectrum magnitude for the hydrophone time series from the 2712 rpm straight-track run (at CPA). The grid spacing is 19.5 Hz which corresponds to the propeller axis rotation frequency.

are quantized to 10 levels and displayed using the upper half of the grayscale on the right. There are several peaks that do not fall on the grid, most notably the ones at (approx) (286, 86) Hz, (335, 166) Hz and (334, 263) Hz. Moreover, these peaks do not fall on the grid corresponding to multiples of half of the engine axis frequency either. Therefore, it is conceivable that these coupled frequencies are sums or differences between multiples of the engine and drive frequencies, possibly generated by quadratic phase coupling. However, further study is needed to determine the nature of these peaks.

## 5. CONCLUSION

Only for low speeds (1830 rpm) is it possible to see engine harmonics in the hydrophone data, despite the presence of such harmonics in the hull data. Thus, the hull does not act as a "projector" for engine vibrations. Instead, the dominating source at medium (2712 rpm) to high speed (3549 rpm) is the drive and at high speed only the drive is visible in the bispectrum from hydrophone data. However, using the skewness it is possible to detect coupled harmonics that are neither strictly engine related nor strictly drive related. The propeller leaves a clear trace in both the bispectrum and skewness for medium speeds, in terms of peaks at 7, 8, and 9 times the propeller axis frequency, which might be useful in determining the propeller configuration.



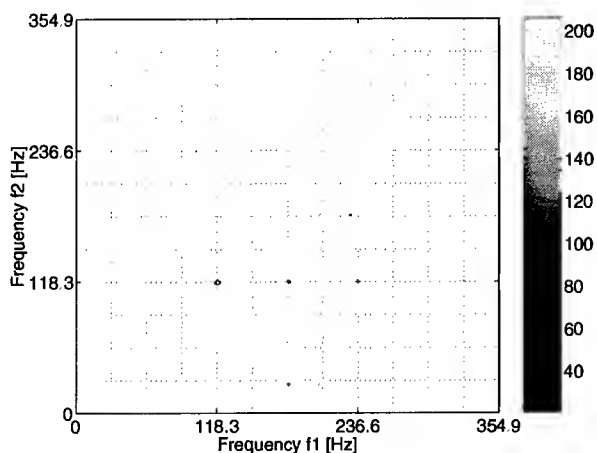


Figure 7: Bispectrum magnitude for the hull accelerometer time series from the 3549 rpm straight-track run (at CPA) with frequency grid corresponding to half the engine axis frequency.

## REFERENCES

- [1] C.L. Nikias and M.R. Raghuver, "Bispectrum Estimation: A Digital Signal Processing Framework," *Proc. IEEE*, Vol. 75, No. 7, pp. 869-891, 1987.
- [2] M. Shen and L. Sun, "The Analysis and Classification of Phonocardiogram Based on Higher-Order Spectra," *Proc. IEEE Workshop on Higher-order statistics*, Banff, Alta., Canada, 21-23 July, 1997, pp. 29-33.
- [3] A.G. Beard, N.J. Mitchell, P.J.S. Williams and M. Kunitake, "Non-linear Interactions Between Tides and Planetary Waves Resulting in Periodic Tidal Variability," *J. Atmosph. Solar Terrest. Phys.*, Vol. 61, pp. 363-376, 1999.
- [4] R.W. Barker and M.J. Hinich, "Statistical Monitoring of Rotating Machinery by Cumulant Spectral Analysis," *Proc. IEEE Workshop on Higher-Order Statistics*, South Lake Tahoe, CA, USA, 7-9 June, 1993, pp. 187-191.
- [5] J.W.A. Fackrell, S. McLaughlin and P.R. White, "Bicoherence Estimation Using the Direct Method. Part 1: Theoretical Considerations," *Applied Sig. Process.*, Vol. 3, pp. 155-168, 1995.
- [6] C.F. Taylor, *The Internal-Combustion Engine in Theory and Practice*, Revised ed., Vol. 2, MIT Press, 1985.

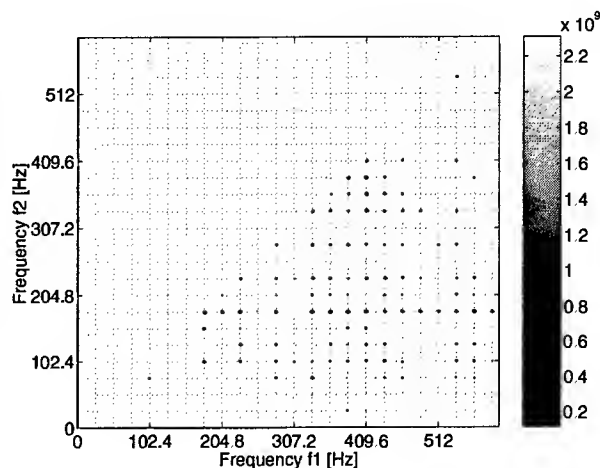


Figure 8: Bispectrum magnitude for the hydrophone time series from the 3549 rpm straight-track run (at CPA) with grid spacing equal to the propeller axis rotation frequency.

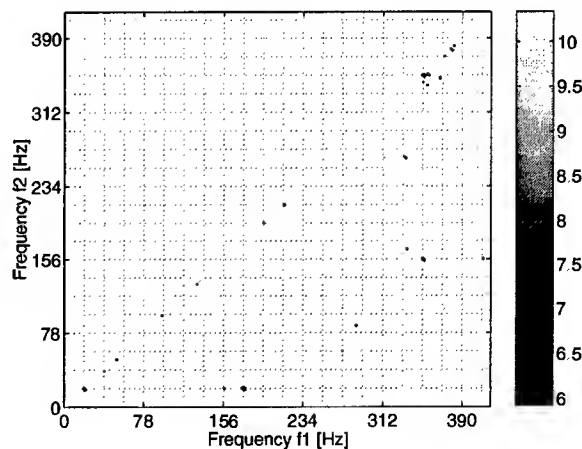


Figure 9: Skewness for the hydrophone time series from the 2712 rpm straight-track run (at CPA). The grid spacing is 19.5 Hz which equals the propeller axis frequency.

# APPROXIMATE CFAR SIGNAL DETECTION IN STRONG LOW RANK NON-GAUSSIAN INTERFERENCE

<sup>†</sup>Ivars P. Kirsteins and <sup>‡</sup>Muralidhar Rangaswamy

<sup>†</sup>Naval Undersea Warfare Center, Code 8212, Newport, RI 02841 USA

<sup>‡</sup>ARCON Corporation, 260 Bear Hill Road, Waltham, MA USA

## ABSTRACT

We have devised a new generalized likelihood ratio test for detecting a signal in unknown, strong non-Gaussian low rank interference plus white Gaussian noise which needs no knowledge of the non-Gaussian distribution. From perturbation expansions of the test statistic, we establish the connection of the proposed GLRT detector to the UMPI test and show that it is approximately CFAR. Computer simulations indicate that the new detector significantly outperforms traditional adaptive methods in non-Gaussian interference.

## 1. INTRODUCTION

Non-Gaussian disturbances have been reported in diverse applications such as radar, sonar, digital communications, and radio astronomy. Signal detection in unknown colored noise backgrounds has traditionally been accomplished using adaptive methods based on the Gaussian model, whether or not the noise is actually Gaussian distributed. However, recent work has shown that the performance of adaptive detectors based on the Gaussian model can degrade severely when operating in correlated non-Gaussian noise backgrounds [1]. To illustrate this, we computer simulated the invariant matched subspace detector (MSD) of Scharf et. al. [2] in noise consisting of a strong, highly correlated rank-2 compound-Gaussian component embedded in white Gaussian noise. Two versions were considered: the optimum MSD that knows the true interference subspace and, motivated by the Principal Component Inverse (PCI) method [3], an adaptive MSD (ASD) that uses an estimate of the interference subspace obtained from signal-free training data. As a reference, we also evaluated the ASD using pure Gaussian noise that had the same nominal covariance matrix as in the non-Gaussian case. The results for all three cases are plotted in figure 1. As is clearly seen, the performance of the ASD degrades substantially in the non-Gaussian noise, whereas, the adaptive detector in pure Gaussian noise has performance nearly identical

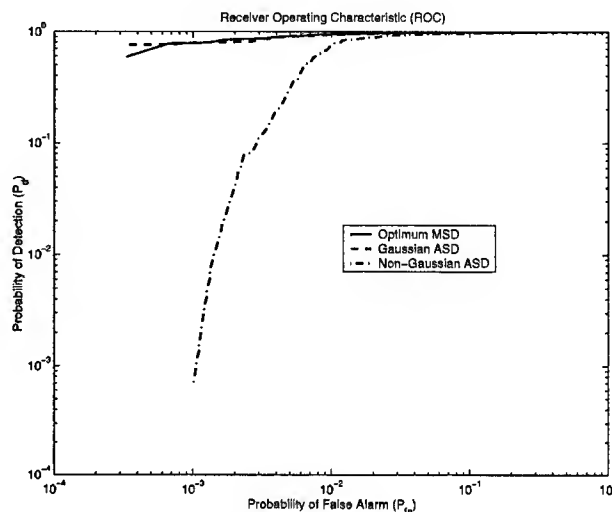


Figure 1: Experimentally measured ROC curves comparing the performance of the detectors at a signal-to-interference ratio of -5 dB.

to the optimum MSD. The effect of non-Gaussian interference on the PCI and subspace methods is discussed in [4].

The underlying problem of designing detectors for non-Gaussian clutter has been the selection of a suitable multivariate probability density function (pdf) family to model the clutter. The difficulty is that in most applications there exists no single family of multivariate non-Gaussian pdfs that accurately characterizes the clutter in all scenarios and environments. Regardless, even if the non-Gaussian pdf family is known, the pdf parameters themselves are usually unknown and their estimation from training data can be problematic. Another difficulty is the sensitivity of parametric pdf estimators and detectors to contaminants in the training data. An alternative approach is to use some sort of non-parametric method, e.g., such as designing locally optimum detectors based on non-parametric kernel-based pdf estimators [5]. However, these methods are best suited for estimating univariate pdfs and are dif-

difficult to extend to higher dimensions and require large amounts of training data.

In many applications where the noise appears to be non-Gaussian, the noise can actually be modeled as consisting of two components: a strong non-Gaussian component which gives rise to the overall non-Gaussian characteristics, and a residual Gaussian part, made up of ambient noise and diffuse clutter. We now propose an alternative approach inspired by the methodology used in [6] to detect weak signals in non-Gaussian Arctic sea noise. Rather than trying to model the overall or individual non-Gaussian characteristics of the noise, a simpler approach is to develop compact representations to model the non-Gaussian and Gaussian waveforms. Then, treating their parameters as unknown, but deterministic, the detection problem can be formulated as a composite hypothesis testing problem [7]. This detection problem is often easier to solve than the original non-Gaussian problem, say by a generalized likelihood ratio test (GLRT).

More precisely we model the received complex-valued  $m \times 1$  noise plus signal space-time data snapshot at time  $t_k$  as a superposition

$$\mathbf{z}_k = \underbrace{\sum_{j=1}^{r_n} a_j^k \mathbf{b}_j}_{\text{subspace interference}} + \underbrace{c^k \mathbf{s}}_{\text{signal}} + \underbrace{\mathbf{n}_k}_{\text{background white noise}} \quad (1)$$

of a strong subspace non-Gaussian interference component and a background white Gaussian noise component  $\mathbf{n}_k$ , and possibly a signal component. The  $a_j^k$  and  $c^k$  are the noise and signal expansion coefficients respectively and the  $\mathbf{b}_j$  and  $\mathbf{s}$  are the noise and signal basis vectors respectively. The non-Gaussianity of the noise is modeled as arising from the expansion coefficients  $a_j^k$  rather than the basis vectors  $\mathbf{b}_j$ . For convenience, a rank-1 signal is assumed.

For the case of known  $\mathbf{b}_j$ , but unknown  $a_j^k$  with unknown multivariate pdf and unknown white noise variance, it is reasonable to seek a test which is invariant to these parameters. Ideally, we desire a uniformly most powerful invariant (UMPI) test [7] (the UMPI test maximizes the probability of detection regardless of the parameter values while keeping the false alarm rate less than or equal to some specified value). Scharf et. al. [2] showed that for data of the form (1) with known interference and signal subspaces, the UMPI test, referred to as the matched subspace detector (MSD), is (in simplified form)

$$\frac{\|P_{P_B^\perp S} \mathbf{z}\|_F^2}{\|P_{BS}^\perp \mathbf{z}\|_F^2} \underset{\mathcal{H}_0}{\overset{\mathcal{H}_1}{>}} \lambda \quad (2)$$

where  $\lambda$  is some threshold. The matrix  $P_{P_B^\perp S}$  is the projection operator onto the part of the signal that remains after the subspace interference has been nulled and  $P_{BS}^\perp$  is the projection operator that nulls out both the subspace interference and signal component. Mathematically,  $P_{P_B^\perp S}$  and  $P_{BS}^\perp$  are given by

$$P_{P_B^\perp S} = P_B^\perp S (S^H P_B^\perp S)^{-1} S^H P_B^\perp \quad (3)$$

and

$$P_{BS}^\perp = I - [B|S]([B|S]^H [B|S])^{-1} [B|S]^H \quad (4)$$

where  $B = [\mathbf{b}_1, \mathbf{b}_2, \dots, \mathbf{b}_{r_n}]$  and  $S = \mathbf{s}$ . The matrix  $[B|S]$  is obtained by concatenating  $B$  and  $S$  column-wise) respectively, and

$$P_B^\perp = I - B(B^H B)^{-1} B^H \quad (5)$$

Test (2) is maximally invariant to scalings of the data and rotations in the column space of  $B$ . Hence it is CFAR with respect to the background noise level. It is emphasized since (2) is UMPI, no other CFAR test can perform better.

Although test (2) is optimum, it is difficult to realize because the interference subspace  $B$  is seldom known beforehand in practice. One approach is to use the methodology of the PCI method [3] and estimate the unknown interference subspace from a set of signal-free training data. However, as the previous and upcoming numerical examples indicate, this approach may not be optimum when the low rank noise is non-Gaussian.

The approach we take is to treat  $B$ ,  $a_j^k$  and the white noise variance as unknown, but deterministic, and derive the GLRT [8] (the GLRT is obtained by replacing the unknown parameters in the likelihood-ratio test by their ML estimates). Our motivation is that in certain instances, the GLRT can actually be UMPI and often leads to a reasonable or good test [2]

## 2. NEW GLRT DETECTOR

A secondary data set of  $K$  signal-free data vectors is assumed available for training, stacked column-wise into a  $m \times K$  matrix  $X$ . Detection of the signal is to be performed on a primary data set, consisting of a single data snapshot, denoted as  $Y$ . Under the null hypotheses  $\mathcal{H}_0$  and signal present hypotheses  $\mathcal{H}_1$ , the observed data matrices  $Z = [X|Y]$  are modeled as

$$\mathcal{H}_0 : Z = BA + N \quad (\text{noise only}) \quad (6)$$

$$\mathcal{H}_1 : Z = BA + [0|Sc] + N \quad (\text{signal + noise}) \quad (7)$$

where  $B$  is a  $m \times r_n$  matrix whose columns generate the low rank interference space,  $A$  is a  $r_n \times K + 1$  matrix whose elements contain the low rank interference expansion coefficients,  $S$  is a  $n \times 1$  signal replica, and  $c$  is the signal amplitude. The elements of matrix  $N$  are modeled as IID complex Gaussian random variables with zero-mean and variance  $\sigma^2$ .  $S$  is assumed known, but  $A$ ,  $B$ ,  $c$ , and  $\sigma^2$  are assumed to be unknown, but deterministic.

A GLRT statistic for the hypothesis testing problem of (6) and (7) is then

$$y \equiv \frac{\max_{B_1, A_1, c, \sigma_1^2} (\sigma_1^2)^{-mK} e^{-\frac{1}{\sigma_1^2} \|Z - B_1 A_1 - [0|Sc]\|_F^2}}{\max_{B_0, A_0, \sigma_0^2} (\sigma_0^2)^{-mK} e^{-\frac{1}{\sigma_0^2} \|Z - B_0 A_0\|_F^2}} \quad (8)$$

which simplifies to the ratio of fitting errors

$$y \equiv \frac{\min_{B_0, A} \|Z - B_0 A\|_F^2}{\min_{B_1, A_1, c} \|Z - B_1 A_1 - [0|Sc]\|_F^2} \quad (9)$$

The numerator of (9) is the square-error in fitting the matrix  $Z$  by a rank  $r_n$  matrix and can be easily evaluated using the SVD of  $Z$ . Similarly, the denominator of (9) is the error in jointly fitting  $Z$  by a rank  $r_n$  matrix and the linear part  $[0|Sc]$ . However, it can not be directly evaluated using the SVD of  $Z$ .

To numerically evaluate the denominator, we propose a simple scheme that is based on a criss-cross regression-like method. The idea is to linearize the minimization by holding, say  $B$ , constant and then minimizing with respect to only  $A$  and  $c$ . This is a standard linear least-squares fitting problem and is easy to solve. The procedure is then repeated, this time replacing  $A$  with its estimate from the previous step and now minimizing with respect to  $B$  and the  $c$ . These steps are repeated until convergence.

### 3. RELATIONSHIP TO UMPI DETECTOR

We now establish the connection of the proposed GLRT to the UMPI matched subspace detector of Scharf et al. [2] by deriving a simple approximation to the test statistic. First, in order to make the comparison, we need to extend the single data vector optimum MSD (2) to the multiple data vector case of (6) and (7). This is simple to do and by substitution (by concatenating all the columns of  $Z$  into one vector), we obtain the optimum MSD test statistic for the multiple data vector case:

$$y_{MSD} - 1 \equiv \frac{\|P_{S'} z'\|_F^2}{\|P_{S'}^\perp z'\|_F^2} \quad (10)$$

where  $z' = \text{vec}(P_B^\perp Z)$ ,  $S' = [\text{vec}(P_B^\perp [0|S])]$ ,  $P_{S'} = S'(S'^H S')^{-1} S'$ , and  $P_{S'}^\perp = I - P_{S'}$ .

We now use a first-order perturbation expansion for the SVD of a data matrix [9] to obtain an approximation to the GLRT test statistic (9) which can be related to the UMPI MSD (10). In the analysis, both  $Sc$  and  $N$  are regarded as perturbations and weak relative to  $BA$ . The specific derivation details are shown in appendix A. The final approximation for the GLRT statistic derived in appendix A is

$$y - 1 \approx \frac{\|P_{S''} z''\|_F^2}{\|P_{S''}^\perp z''\|_F^2} \quad (11)$$

where  $z'' = \text{vec}(P_B^\perp Z P_A^\perp)$ ,  $S'' = \text{vec}(P_B^\perp [0|S] P_A^\perp)$ ,  $P_{S''} = S''(S''^H S'')^{-1} S''$ ,  $P_{S''}^\perp = I - P_{S''}$ , and  $P_A^\perp = I - A^H(AA^H)^{-1}A$ .

The only difference between the UMPI MSD (10) and the new GLRT (11) is the post multiplication of the data matrix  $Z$  by  $P_A^\perp$ . Thus to first-order, the new GLRT is approximately equivalent to the optimum MSD. By inspection, it is seen that (11) is invariant with respect to common scalings of the columns of the data matrix  $Z$ , and thus the background noise level. Thus, the new GLRT is at least approximately CFAR with respect to the background noise level.

When the interference is strong and signal weak, the loss in performance of the GLRT comes from the additional nulling due to the post-multiplication of the data matrix  $Z$  by  $P_B^\perp$ . This loss can be interpreted as arising from having to estimate the interference subspace and is a function of the orthogonality of the interference matrix row space to the row space of the signal matrix  $[0|Sc]$ .

### 4. NUMERICAL EXAMPLES

We now present a numerical example where a 20 element array is used to detect a weak monochromatic signal embedded in strong, highly correlated rank-2 compound-Gaussian clutter plus white Gaussian noise. The output from the array elements is assumed to be already in complex envelope form, so all the data here is complex-valued.

The interference components were computer synthesized as follows: The rank-2 clutter component was modeled as the scattering arising from two independent random discrete reflectors excited by a monochromatic signal pulse located  $\pm 1/2$  DFT bin in wavenumber space symmetrically about broadside. Their amplitude was modeled as a unit variance K-distributed random variable with a shape parameter of .1. Choosing .1 as the shape parameter makes the amplitude distribution heavy-tailed. The background noise samples were modeled as independent and identically distributed zero-mean complex Gaussian random variables.

A total of 24 signal-free data snapshots were used for the secondary or training data set. The primary data set for detection consisted of a single data snapshot. The white noise variance was set to .1 giving a interference-to-white-noise ratio of 10 dB. The signal direction of arrival was chosen to be broadside to the array and the signal power to interference ratio ( $10\log_{10}\sigma^2$ ) was set to -5 dB. 15000 independent trials with and without a signal were performed, computer simulating the new GLRT, optimum MSD, ASD, and Kelly's CFAR GLRT [10] receivers. For comparison, an analogous pure Gaussian noise case with the same nominal covariance matrix was also simulated. Note that the ASD was implemented by using the 24 snapshot signal-free secondary data set to estimate the rank-2 interference subspace via a SVD and plugging the estimated noise subspace into (2).

Figures 2 and 3 show the empirically measured probability of detection (pd) curves obtained for a probability of false alarm (pfa) of .005 for the non-Gaussian and Gaussian cases respectively for all four detectors. From the pd curves in figure 2, it can be seen that the new GLRT has nearly the same performance as the optimum MSD and significantly outperforms the ASD and Kelly's GLRT when the interference is compound-Gaussian. However, it is interesting to observe that for the pure Gaussian case (figure 3), both the new GLRT and the ASD perform almost as well as the optimum MSD.

One last question to be resolved is the degree to which the new GLRT statistic distribution under the null hypotheses is affected by the distribution of the low rank interference component. The perturbation analysis approximation (11) suggests that the GLRT is CFAR to at least first-order. However, the analysis ignores any higher order terms. To obtain insight, we computer simulated the new GLRT using the previous non-Gaussian and Gaussian example for 20000 independent trials for the null hypotheses only. We then calculated the empirical cumulative distribution function of the test statistic and used it to determine the threshold to achieve a given pfa. Figure 4 shows the pfa plotted as a function of threshold for both the non-Gaussian and Gaussian cases. As can be seen from figure 4, the pfas are very close. The pfas only slightly deviate as the threshold increases, implying that the new GLRT is approximately invariant to the distribution of the low rank interference component.

## 5. CONCLUSION

We have derived a new GLRT detector and shown its relationship to the UMPI MSD. Our perturbation

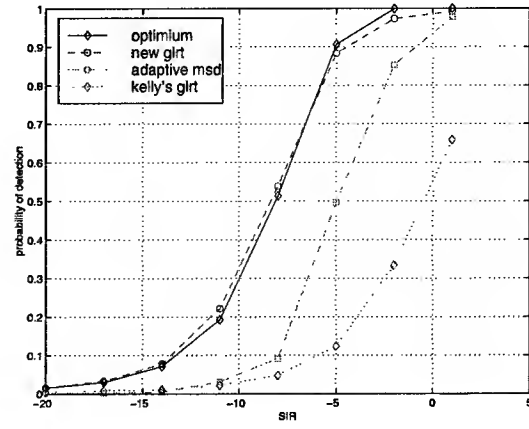


Figure 2: Experimentally measured probability of detection in non-Gaussian interference at a pfa of .005 based on 15000 trials.

analysis and numerical examples show that the new GLRT is likely to be much more robust in low rank non-Gaussian clutter than *ad hoc* or conventional adaptive detectors. Finally, further work needs to be done in analyzing the detectors performance in regards to signal and rank mismatch and higher-order effects due to the non-Gaussianity of the interference.

## APPENDIX A: PERTURBATION ANALYSIS

We start with the numerator of (9). Recall that the numerator is the square-error in fitting a rank  $r_n$  matrix to  $Z$ . Letting  $Z = AB + N$ , where  $N$  is some perturbation and using the first-order subspace perturbation expansion derived in [9] for the error in approximating a matrix by a matrix of lower rank, we obtain

$$\min_{B,A} \|Z - BA\|_F^2 \approx \widehat{num} = \|P_B^\perp Z P_A^\perp\|_F^2 \quad (12)$$

where  $P_A^\perp = I - A(A^H A)^{-1} A^H$ .

We now approximate the denominator. If the denominator of (9) is solved with respect to only  $B_1$  and  $A_1$  (holding  $c$  fixed), it is equivalent to finding the rank  $r_n$  approximation to  $Z - [0|S]c$ . Treating  $[0|S]c$  as a perturbation (weak signal and noise case) initially and applying (12), we can approximate the denominator as

$$\widehat{den} \approx \min_c \|P_B^\perp Z P_A^\perp - c P_B^\perp [0|S] P_A^\perp\|_F^2 \quad (13)$$

The minimization of (13) is a standard linear least-squares problem and the residual fitting error is

$$\widehat{den} \approx \|P_{S''}^\perp z''\|_F^2 \quad (14)$$

where  $z'' = \text{vec}(P_B^\perp Z P_A^\perp)$ ,  $P_{S''}^\perp = I - P_{S''}$ ,  $P_{S''} = S''(S''^H S'')^{-1} S''^H$ , and  $S'' = [\text{vec}(P_B^\perp [0|S] P_A^\perp)]$ . The

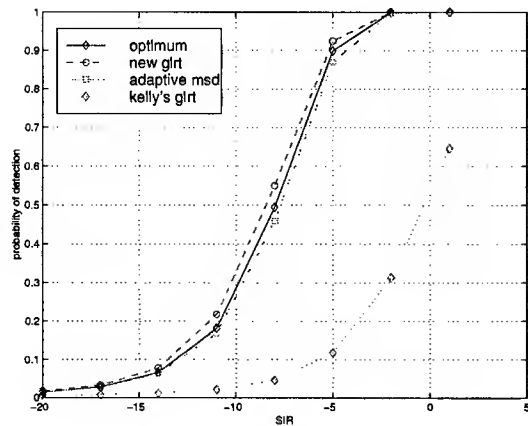


Figure 3: Experimentally measured probability of detection in Gaussian interference at a pfa of .005 based on 15000 trials.

operator  $\text{vec}(\cdot)$  takes a matrix and converts it to a vector representation by stacking the columns. Finally, replacing the exact quantities in (9) by their above approximations (12) and (14), and after some simplification, we obtain

$$y \approx 1 + \frac{\|P_{S''} z''\|_F^2}{\|P_{S''} z''\|_F^2} = \frac{\widehat{\text{num}}}{\widehat{\text{den}}} \quad (15)$$

## REFERENCES

- [1] M. Rangaswamy, and J.H. Michels, "A parametric detection algorithm for space-time processing in non-Gaussian clutter," *Signal Processing*, 1999 (to appear).
- [2] L.L. Scharf and B. Friedlander, "Matched subspace detectors," *IEEE Trans.* vol. 42, no. 8, pp.2146-2156, Aug. 1994.
- [3] I.P. Kirsteins and D.W. Tufts, "Adaptive detection using low rank approximation to a data matrix," *IEEE Trans. AES*, 30(1):55-67, Jan. 1994.
- [4] M. Rangaswamy, B.E. Freburger, I.P. Kirsteins, and D.W. Tufts "Signal Detection in Strong Low Rank Compound-Gaussian Interference," in *Proc. of the IEEE SAM2000 workshop*, Boston, MA, March 2000.
- [5] S.M. Zabin and G.A. Wright, "Nonparametric density estimation and detection in impulsive interference channels - part II: Detectors," *IEEE Trans. on Comm. Theory*, vol. 42, no. 2/3/4, p-p. 1698-1711, Feb./March/April 1994.

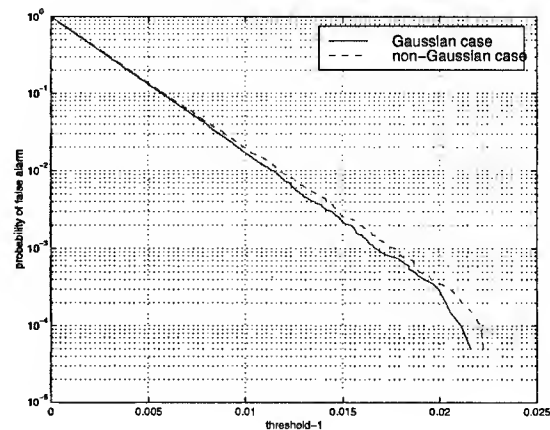


Figure 4: Experimentally measured probability of false alarm plotted as a function of threshold for both the Gaussian and non-Gaussian cases.

- [6] I.P. Kirsteins and D.W. Tufts, *Methods of computer-aided analysis of non-Gaussian noise and application to robust adaptive detection: Part 2*, DTIC report ADA148879, September 1984.
- [7] T.S. Ferguson, *Mathematical Statistics: A Decision Theoretic Approach*, Academic Press, New York, 1967.
- [8] H.L. Van Trees, *Detection, Estimation, and Modulation Theory*, John Wiley and Sons, 1968.
- [9] R.J. Vaccaro, "A second order perturbation expansion for the SVD," *SIAM J. Matrix Anal. Appl.*, 15(2):661-671, April 1994.
- [10] E.J. Kelly, "An Adaptive Detection Algorithm," *IEEE Trans. on Aerospace and Electronic Sys.*, Vol. AES-23, No. 4, pp. 115-127, Nov. 1986.

# BLIND EQUALIZATION OF PHASE ABERRATIONS IN COHERENT IMAGING: MEDICAL ULTRASOUND AND SAR

Seth D. Silverstein

University of Virginia, Department of Electrical Engineering  
Charlottesville, VA 22903  
silverstein@virginia.edu

## ABSTRACT

This work introduces a novel auto-focusing phase aberration correction algorithm for coherent imaging systems such as medical ultrasound and synthetic aperture radar (SAR). The algorithm follows directly from the analytical expressions obtained from a detailed theoretical analysis of the scattering of wave packets in a random scattering medium consisting of many scatterers in each resolution cell. The algorithm selects the resolution cell in a range-azimuth region with the smallest ratio of the variance of the focused channel/element amplitudes to the mean square of the focused channel amplitudes. The aberrant phase can be estimated directly from this selected cell using well known phase gradient techniques. *Monte Carlo* simulations are shown to exhibit excellent algorithmic performance.

## 1. INTRODUCTION

In this work we present a novel auto-focusing phase aberration correction algorithm for coherent imaging systems such as medical ultrasound and synthetic aperture radar (SAR). The algorithm follows a detailed theoretical analysis of the scattering of wave packets in a random scattering medium consisting of many scatterers in each resolution cell. Phase aberration detection and correction algorithms have been studied extensively in the literature[1-5]. Some degree of success has been achieved for the radar imaging applications, however the medical ultrasound algorithms have exhibited limited success in actual *in vivo* ultrasound experiments.

The propagation velocity of the signals in medical ultrasound and Synthetic Aperture Radar differ by five orders of magnitude. Nonetheless, both of these signals satisfy a similar wave equation, and there are significant commonalities in the physics based signal processing of the coherent imaging processes. High resolution images are of great importance in both these modalities. In medical diagnostic ultrasound, one desires, for example, to distinguish a cancerous lesion from a cyst. High resolution surveillance SAR systems are used to automatically characterize targets. In both modalities phase errors in the image formation cause blurring which in turn reduces the efficacy of the characterization algorithms.

A phased array coherent imaging system is focused at a particular resolution cell by adjusting the phases received at all the elements of the array such that an interference maximum occurs for the coherent signals scattered from the focused cell. As we scan throughout the image space, the strength of the scattering from the focused cells will form a brightness image of the scattering field. In a digital coherent imaging system, the received signals are sampled at each of the elements of the spatially sampled coherent aperture. The focused image map of a resolution cell is obtained by selecting the time samples at all the array elements such that the time of flight, and hence the phase, of the coherent scattered signals from the resolution cell of interest *will be the same*.

The phase in corrupt data signals is a combination of the necessary *good* phase that contains the geometrical and scattering phase information that would occur in a non-corrupted scattered signal, and the aberrant *bad* phase that serves to blur the image. The problem that needs to be solved is to develop an algorithmic that provide some means of differentiating the good from bad phase. Once this differentiation can be established, the bad phase can be estimated and selectively pruned out of the signals. The *cleaned* signals will retain enough of the good phase so the net effect of the cleaning procedure on the image quality is positive.

As stated above, focusing in digital medical ultrasound sound systems is accomplished by choosing the delays according to estimates propagation times based upon estimated sound velocities. Human tissue has sub-cutaneous layers of fatty tissue that are not observable from the surface. The sound velocity difference between the fatty and muscle tissue can cause a significant error in the time delay focusing which in turn causes phase errors and an associated blurring of the image.

In SAR the coherent aperture is synthetically generated as a T/R physical antenna is flown over an extended flight path. Coherent pulses are transmitted and the scattered signals are subsequently received at points in the flight path separated by the velocity of the vehicle divided by the coherent pulse repetition frequency. In the SAR scenario we are dealing with, for example, X-band radars with 3 cm wavelengths. Small, UN-compensated deviations in the flight path amounting to a fraction of a wavelength over flight paths of the order of a kilometer can cause significant blurring. Here again, we can apply a slight variation of the physics based signal processing algorithm that we will de-

This work was supported by NSF under grant #CCR-9817630

tail for ultrasound to blindly estimate and correct for the SAR phase aberrations.

The enabling mechanism that makes this algorithm perform so well follows directly from the mathematical analysis of the scattering relations. We have shown that the variance of the time selected-focused elemental(channel) signals depends strongly upon the azimuthal distribution of the scatterers in the focused resolution/scan cell. If the resolution cell scatterers are narrowly distributed in azimuth around the focused angle, the variance of the elemental signal amplitudes across the array will be small. The algorithm therefore searches for the resolution cell with the smallest variance. It is important to normalize the variance by dividing by the square of the mean of the channel amplitudes. This normalization scales out the effects of overall brightness from the measure, as possible errors in the Minimum Variance Scan Cell (MVSC) assignment could be made for a very dim speckle in the absence of the normalization. As the part of the good phase that is element dependent will be very small for the MVSC, we can effectively assume that the good phase has been stripped off the channel signals associated with the MVSC. The estimate of the remaining bad, channel-dependent-phase is made using well known phase gradient techniques.

In the subsequent sections of this paper we restrict discussion to the ultrasound application.

## 2. MODEL OF SCATTERED SIGNALS

In a ultrasound polar B scan the transmitted signals associated with a single firing of the elements of the array are individually time delayed to produce interference maxima at transmit foci located at specific ranges and azimuthal scan angles. On receive, the echo time-sampled coherent signals are first accumulated at each element. The focus-on-receive signals are then constructed from the coherent sum of the individual elemental time samples that are chosen according to their propagation times from the receive foci to the individual elements. The B scan receive foci are at the same azimuthal angles as the transmit foci, but at different ranges. The received-sampled B scan elemental signals have three indices, an element index,  $n$ , a range index,  $k$ , and an azimuthal scan index  $\ell$ . Here the number of elements(channels) is given by  $N_a$ . The corrupted signal can be generally represented by,

$$s_e(n, k, \ell) = e^{j\phi_{k\ell}(n)} s(n, k, \ell). \quad (1)$$

The auto-focusing method seeks to obtain an estimate of the aberrant phases,  $\hat{\phi}_{k\ell}(n)$ . Then, an estimate of the *cleaned* signal is constructed by equalizing the effects of the aberrant phase,

$$\hat{s}_c(n, k, \ell) = e^{-j\hat{\phi}_{k\ell}(n)} s_e(n, k, \ell). \quad (2)$$

The model can be simplified by dividing the image map up into azimuthal and/or range segments. One can assume that the aberrant phases are weakly dependent upon either the range or scan angle indices within the sector domains. In the examples described herein we retain a range dependence on the aberrant phase as we will process all the range cells individually. In an alternative procedure which is computationally less intensive, the range cell dependence

is neglected over a range segment and the aberrant phase estimated for a single range cell is used for all the range cells in the segment. Our examples assume:

$$\phi_{k\ell}(n) \approx \phi_k(n), \quad \text{in azimuthal segment.}$$

The transmit foci coordinates  $\vec{R}_{ft}(\ell)$  are characterized by a range,  $R_{ft}$ , and azimuthal bearing angle,

$$\mu_{ft}(\ell) \equiv \sin \theta_{ft}(\ell) = \frac{\vec{r}_n \cdot \vec{R}_{ft}(\ell)}{r_n R_{ft}}. \quad (3)$$

The origin of the coordinate system is set at the center of the array. The sample scan angles are represented by

$$\mu_{ft}(\ell) = \mu_0 + [\ell - (N_\ell + 1)/2] \Delta_{\mu\text{scan}}; \quad \ell = 1, 2, \dots, N_\ell \quad (4)$$

Here  $N_\ell$  is the number of B scan lines in the B scan segment. and  $\mu_0$  is the central azimuth of the azimuthal scan. The azimuthal partition,  $\Delta_{\mu\text{scan}}$  is typically taken to be a fraction  $\sim 1/4$  to  $1/2$  of the Rayleigh angular resolution of the array,  $\sin(\lambda_c/L)$ . Here  $\lambda_c$  is the central wavelength of the transmitted ultrasound pulse, and  $L$  is the length of the coherent aperture of the ultrasonic probe.

The B scan receive foci have the vector coordinates,  $\vec{R}_{fr}(k, \ell)$ , with magnitudes,  $R_{fr}(k)$ , and angle,  $\mu_{fr}(\ell) \equiv \mu_{ft}(\ell)$ . For notational simplicity we define

$$\mu_f(\ell) \equiv \mu_{fr}(\ell) \equiv \mu_{ft}(\ell).$$

The range cell partition of the receive foci is represented as

$$R_{fr}(k) = R_{ft} + \Delta(k); \quad \Delta(k) \stackrel{\text{def}}{=} [(k-1) - (k_{\text{max}}-1)/2] \Delta_R; \quad (5)$$

for  $k = 1, 2, \dots, k_{\text{max}}$ . Typical values for the range partition,  $\Delta_R$ , are taken to be  $\sim 1/4$  to  $1/2$  of range resolution associated with the bandwidth,  $c_s/(2B)$ . Here  $c_s$  is the speed of sound, and  $B$  is the bandwidth of the ultrasonic pulse.

A general formulation of the scattering of ultrasound can be very complex if one tried to model multiple scattering effects. The scattering equations become tractable when the *Born* approximation is invoked. The Born model considers the effects of single scattering events only. Here the received signal just prior to sampling at the receiving element at  $\vec{r}_n$  that is scattered from a point scatterer at  $\vec{r}_s$  after being initially transmitted from a transducer centered at  $\vec{r}_m$  is modeled as

$$s(\vec{r}_t, \vec{r}_s, \vec{r}_m, t) = \frac{A_s}{(4\pi)^2} \frac{\alpha(\vec{r}_s, \vec{r}_m) \alpha(\vec{r}_s, \vec{r}_t)}{|\vec{r}_s - \vec{r}_m| |\vec{r}_s - \vec{r}_t|} \times \int e^{j\omega(t - (|\vec{r}_s - \vec{r}_m| + |\vec{r}_s - \vec{r}_t|)/c_s)} K_0^2(\omega) \frac{d\omega}{2\pi}. \quad (6)$$

Here  $K_0(\omega)$  is the Fourier transform of the impulse response of the transmitter/receiver elements, and  $\alpha(\vec{r}_s, \vec{r}_n)$  represents the transducer element factor. For the considerations presented herein we use a simple cosine obliquity factor for the element factor,

$$\alpha(\vec{r}_s, \vec{r}_n) \stackrel{\text{def}}{=} \frac{(\vec{r}_s - \vec{r}_n) \cdot \vec{n}_\perp}{|\vec{r}_s - \vec{r}_n|}. \quad (7)$$



The mathematics formulation of the propagation and scattering of ultrasonic signals of the type used in medical diagnostic applications can be simplified using a spherical Gaussian model for the propagating wavepackets. A Gaussian wave form is particularly simple to work with analytically as the Fourier transform of a Gaussian is also a Gaussian. Consider a normalized Gaussian pulse  $K_0(t)$  that is centered at the transducer at the time  $t = 0$ . The pulse and its Fourier transform  $K_0(\omega)$  are represented by,

$$K_0(t) = \frac{1}{\sqrt{2\pi\tau^2}} e^{j\omega_c t} e^{-\frac{t^2}{(2\tau^2)}} \xrightarrow{\text{trans}} K_0(\omega) = e^{-\frac{(\omega - \omega_c)^2}{(2B^2)}} \quad (8)$$

Here the bandwidth  $B \equiv 1/\tau$ . The fractional bandwidths<sup>1</sup> of interest fall into a broad range from  $\sim 20 - 80\%$ . Using a Gaussian pulse form, the Fourier integral on the right-hand-side of Eq. (5) can be readily evaluated,

$$\int e^{j\omega T_{ret}(n,s,m)} K_0^2(\omega) \frac{d\omega}{2\pi} = \frac{B}{\sqrt{4\pi}} e^{j\omega_c T_{ret}(n,s,m)} e^{-B^2 T_{ret}^2(n,s,m)/4} \quad (9)$$

We have used an abbreviated notation for the retarded time,

$$T_{ret}(n,s,m) = t - t_{0m} - [|\vec{r}_s - \vec{r}_n| - |\vec{r}_s - \vec{r}_m|]/c_s, \quad (10)$$

where  $t_{0m}$  is the initiation time of the pulse at the  $m$ th transmitting element. Summing over all scatterers, we have

$$s(\vec{r}_n, \vec{r}_m, t) = C \sum_s A_i \frac{\alpha(\vec{r}_s, \vec{r}_n) \alpha(\vec{r}_s, \vec{r}_m)}{|\vec{r}_s - \vec{r}_n| |\vec{r}_s - \vec{r}_m|} \times e^{j\omega_c T_{ret}(n,s,m)} e^{-B^2 T_{ret}^2(n,s,m)/4} \quad (11)$$

Here  $C$  is a generic constant that captures all the constant coefficients and factors of  $\pi$  that arise in the process.

In time delay focusing the time delays of the transmitted signals in a single firing of the array cause all the coherent pulses to arrive at the transmit focus at the same time represented by  $t_{ft}(\ell)$ ,

$$t_{ft}(\ell) = \{t_{0m} + |\vec{R}_{ft}(\ell) - \vec{r}_m|/c_s\}, \text{ for all } m. \quad (12)$$

The samples of the scattered received signals are then chosen so that their sample delay times corresponds to an interference maximum from a scatterer located at the receive focus  $\vec{R}_{fr}(k, \ell)$ . This corresponds to the receive time samples,

$$t_n(k, \ell) = t_{ft}(\ell) + [|\vec{R}_{ft}(\ell) - \vec{r}_n| + |\vec{R}_{fr}(k, \ell) - \vec{R}_{ft}(\ell)|]/c_s, \quad (13)$$

at the  $n$ th receiver. The retarded times corresponding to the B scan scattered signals that are focused on both transmit and receive are,

$$T_{fret}(n, s, m, k, \ell) = \frac{1}{c_s} \left[ \Delta(k) + |\vec{R}_{ft}(\ell) - \vec{r}_m| + |\vec{R}_{fr}(k, \ell) - \vec{r}_n| - |\vec{r}_s - \vec{r}_n| - |\vec{r}_s - \vec{r}_m| \right]. \quad (14)$$

<sup>1</sup>For a Gaussian pulse the bandwidth is defined here as twice the 3dB width of  $K_0^2(\omega)$ . The fractional bandwidth  $F$ , and bandwidth  $B$  have the functional relationship,  $B = \frac{1}{2\sqrt{\ln 2}} F \omega_c$

Here the  $\{\vec{r}_s\}$  are the coordinates of the scatterers. The received signals that are focused at  $\vec{R}_{fr}(k, \ell)$  are

$$s(n, k, \ell) = \sum_s A_s \gamma(n, s, k, \ell) \frac{\alpha(\vec{r}_n, \vec{r}_s)}{|\vec{r}_n - \vec{r}_s|} \times e^{j\omega_c [|\vec{R}_{fr}(k, \ell) - \vec{r}_n| - |\vec{r}_s - \vec{r}_n|]/c_s} \quad (15)$$

The function  $\gamma(n, s, k, \ell)$ ,

$$\gamma(n, s, k, \ell) \stackrel{\text{def}}{=} C e^{j\omega_c \Delta(k)/c_s} \sum_m |w_m| \frac{\alpha(\vec{r}_m, \vec{r}_s)}{|\vec{r}_m - \vec{r}_s|} \times e^{-B^2 T_{fret}(n, s, m, k, \ell)} e^{j\omega_c [|\vec{R}_{ft}(\ell) - \vec{r}_m| - |\vec{r}_s - \vec{r}_m|]/c_s}, \quad (16)$$

will be a relatively slowly varying function of the receiver indices  $\{n\}$ . From Eqs. (15,16) we see that the major contributions to the scattered signals will come from scatterers that are in close proximity to the received focus corresponding to values of  $T_{fret}^2(n, s, m, k, \ell) \leq 4/B^2$ . A dominant scatterer will manifest itself as a maximum in the  $B$  scan at a bearing angle in the neighborhood of the central bearing angle of the scattering cluster.

### 3. MVSC ALGORITHM

The first step in the algorithmic procedure is an *Amplitude equalization* to mitigate possible distortion effects caused by transducer obliquity factors and the  $1/R$  dependence of the received signals in the near field.

**Step 1. Amplitude equalization** – the received signal amplitudes are equalized as follows:

$$\tilde{s}_e(n, k, \ell) \stackrel{\text{def}}{=} \frac{|\vec{R}_{fr}(k, \ell) - \vec{r}_n|}{\alpha(\vec{R}_{fr}(k, \ell), \vec{r}_n)} s_e(n, k, \ell). \quad (17)$$

Here  $s_e(n, k, \ell)$  are the corrupt data modeled by Eq. (1). Defining  $\tilde{\gamma}$  as the amplitude equalized form of  $\gamma$ , the individual amplitude equalized corrupt signals are of the form,

$$\tilde{s}_e(n, k, \ell) = \sum_s \tilde{\gamma}(n, s, k, \ell) e^{j\omega_c [|\vec{R}_{fr}(k, \ell) - \vec{r}_n| - |\vec{r}_s - \vec{r}_n|]/c_s} e^{j\phi_k(n)}. \quad (18)$$

We want to estimate the aberrant phase,  $\phi_k(n)$ , given the measured corrupt equalized signal  $\tilde{s}_e(n, k, \ell)$ .

**Step 2.** The second step accumulates and stores the ratio of the variance to square of the mean of the equalized elemental channel signals associated with each of the range-azimuth cells in the scan. This parameter represents the *normalized cell variance* (NV) metric. The NV can be mathematically represented by

$$G(k, \mu_f(\ell)) \stackrel{\text{def}}{=} \frac{N_a \sum_{n=1}^{N_a} |\tilde{s}_e(n, k, \ell)|^2}{\left[ \sum_{n=1}^{N_a} |\tilde{s}_e(n, k, \ell)| \right]^2} - 1. \quad (19)$$

The next processing stage sorts the NV's, and identify the specific cell with the *smallest* NV within the sector scan. This cell is the *minimum variance scan cell* (MVSC). The algorithm then performs the computations that estimate

the aberrant phase from the complex channel signals associated with the MVSC.

The estimation of the phase can be simply understood by examining the approximate functional form of the signals in the Fresnel zone of the array. Here we have,

$$\tilde{s}_e(n, k, \ell) \cong e^{j\phi_k(n)} \sum_s \Gamma(s, k, \ell) \times e^{-j\omega_c r_n(\mu_f(\ell) - \mu_s)[1 + r_n(\mu_f(\ell) + \mu_s)/(2R_{fr}(k))]} \quad (20)$$

Therefore the channel amplitude will have the approximate form.

$$\left| \sum_s \Gamma(s, k, \ell) e^{-j\omega_c r_n(\mu_f(\ell) - \mu_s)[1 + r_n(\mu_f(\ell) + \mu_s)/(2R_{fr}(k))]} \right|. \quad (21)$$

The variance of the channel signals for the focused pixel depends strongly upon the azimuthal distribution of the scatterers in the focused pixel. If the pixel scatterers are narrowly distributed in azimuth around  $\mu_f(\ell)$ , the variance of the channel signal amplitudes across the array will be small. This metric further normalizes the variance by dividing by the square of the mean of the channel amplitudes. This normalization scales out the effects of overall brightness from the measure, as possible errors in the MVP assignment could occur for an inappropriate, very dim speckle, in the absence of the normalization. As the part of the good phase that is channel dependent, *will be very small* for the MVSC, we can effectively assume that the good phase has been stripped off the channel signals associated with the MVSC, i.e.,

$$e^{j\omega_c r_n(\mu_f(\ell) - \mu_s)[1 + r_n(\mu_f(\ell) + \mu_s)/(2R_{fr}(k))]} / c_s \sim 1. \quad (22)$$

The estimate of the remaining bad, channel dependent, phase is made using a phase gradient technique that is well known in the art.

**Step 3. Estimation of aberrant phase** – Now that the estimate of the good phase has been stripped off the signal, the remaining bad channel dependent phase can be estimated using a phase gradient technique,

$$\widehat{\Delta\phi}_k(n) = \angle \left( \tilde{s}_e^*(n-1, k, \ell_{\min}) \tilde{s}_e(n, k, \ell_{\min}) \right). \quad (23)$$

The estimate the phase aberration across the aperture is obtained by integrating the estimated gradient,

$$\hat{\phi}_k(n) = \sum_{q=2}^n \widehat{\Delta\phi}_k(q). \quad (24)$$

**Step 4. Construct estimate of corrected data** – Go back to the original corrupt data set and strip off the estimate of the aberrant phase.

$$\hat{s}(n, k, \ell) = e^{-j\hat{\phi}_k(n)} s_e(n, k, \ell). \quad (25)$$

**Step 5. Subsequent iterations**

The estimation of the aberrant phase can further improved by going back to step 1. for additional iterations of the algorithm.

## 4. SIMULATIONS

Simulation model parameters

$f_c = \frac{\omega_c}{2\pi} = 5 \times 10^6 \text{ Hz.}$	– pulse central frequency
$c_s = 1.54 \times 10^3 \text{ m/sec}$	– sound velocity
$\lambda_c = 3.08 \times 10^{-3} \text{ m}$	– pulse central wavelength
$\Delta f_c = .6 f_c \text{ Hz}$	– modulation bandwidth
$d = \lambda_c / 2$	– trans/receiver element spacing
$B = .850 \pi \Delta f_c \text{ radians/sec}$	– 1/2 mod bandwidth @ 3db.
$Na = 64$	– # T/R elements
$L = (Na - 1)\lambda_c / 2$	– array length
$R_{ft} = 1.5L$	– transmit focus distance.
$f / = 1.5$	– f # of scattering cell
$Vt = 1$	– scat. strength

We simulate the B scan scattering pattern for a randomly generated distribution of point scatterers with approximately 50 scatterers in each of the range/azimuthal resolution cells. This is a sufficient number to generate a fully developed speckle pattern. We have also included a void region of azimuthal width corresponding to  $\sim 1.5$  Rayleigh azimuthal resolution units. Two different random distributions of scatterers with voids are illustrated in Figs. 2,3. The dotted rectangles in the figures illustrate the position resolution cell associated with MVSC cell for these phantoms.

We now impose a random aberration phase. The aberration phase is constructed as follows. Take a sequence  $N_r$  of uniformly distributed random numbers over the interval  $(-0.5, 0.5)$ . Take FFT, low pass filter by using the first  $k_c$  values setting the rest of the transform coefficients equal to zero. Then, take the IFFT and scale the results so the resultant phases amplitudes  $\in \pm\pi$ . This procedure introduces an element to element correlation length of the order of  $(N_r / k_{cu})\lambda / 2$  into the random aberrant phase.

Figs. 3,4 illustrate the results of the first, second, and third iterations of our methodology of blindly estimating and cleaning the results of the aberrant phase for the respective distributions shown in Figs. 1,2. In each of the examples 20 independent random trials of the phase aberration and subsequent cleaning process are illustrated. The results are shown for aberrant phase with a correlation length of 1.38mm. Simulation results over a broad range of aberrant phase correlation lengths demonstrate that our auto-focusing algorithm is effective down to correlation lengths  $\sim 4$  wavelengths for phase aberrations varying between  $\pm\pi$  over the length of the array.

## 5. REFERENCES

1. B. D. Steinberg, "Radar imaging from distorted arrays: the radio camera algorithm and experiments," *IEEE Trans. Antennas Propagat.*, vol. 29, no. 5, pp. 740-748, September 1981
2. B. D. Steinberg, "Microwave imaging of aircraft," *Proc. IEEE*, vol. 76, no. 12, pp. 1578-1592, December 1988.
3. E. H. Attia and B. D. Steinberg, "Self-coherent large antenna arrays using the spatial correlation properties of radar clutter," *IEEE Trans. Antennas Propagat.*, vol. 37, no. 1, pp. 30-38, January 1989

4. D. E. Wahl, P. H. Eichel, D. C. Ghiglia, and C. V. Jakowatz, Jr., "Phase gradient autofocus - a robust tool for high resolution SAR phase correction," *IEEE Transactions on Aerospace and Electronic Systems*, Vol. 30, No. 3 pp. 827-835, July 1994.
5. S. W. Flax and M. O'Donnell, "Phase-aberration correction using signals from point reflectors and diffuse scatterers: basic principles," *IEEE Trans. Ultrason., Ferroelec., Freq. Contr.*, vol 35, no. 6, pp. 758-767, November, 1988.

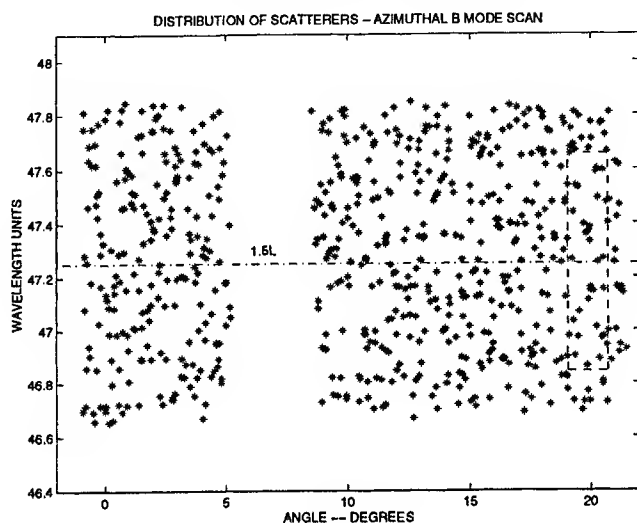


Figure 1: Distribution of scatterers used in simulations

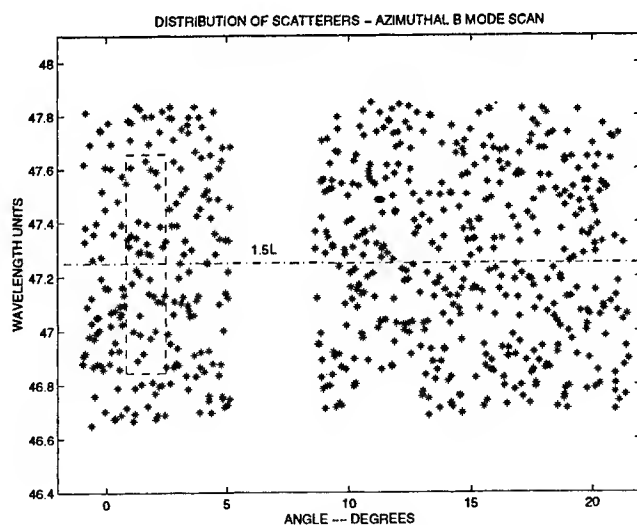


Figure 2: Distribution of scatterers used in simulations

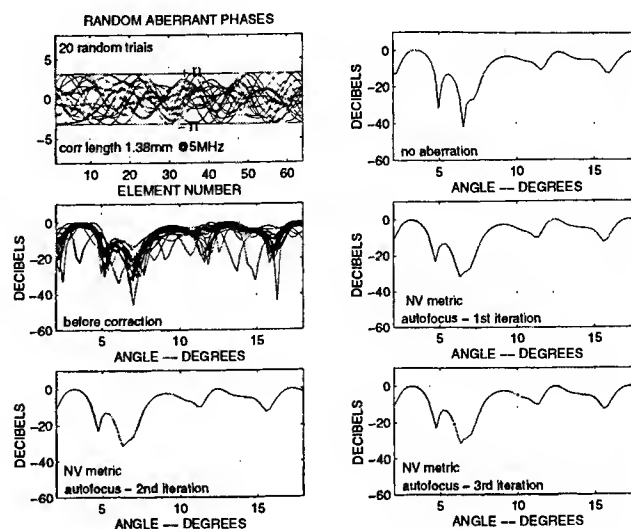


Figure 3: Simulation results for the cleaned images using NVC metric for phantom in Fig. 1.

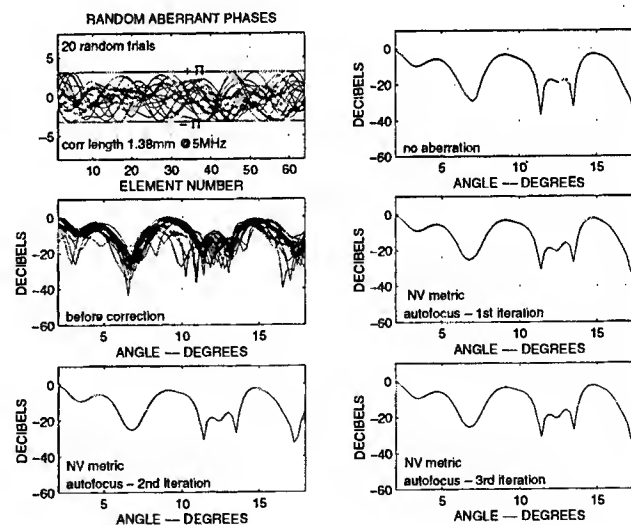


Figure 4: Simulation results for the cleaned images using NVC metric for phantom in Fig. 2.

# FALSE DETECTION OF CHAOTIC BEHAVIOUR IN THE STOCHASTIC COMPOUND K-DISTRIBUTION MODEL OF RADAR SEA CLUTTER

*C.P. Unsworth, M.R. Cowper, S. McLaughlin, B. Mulgrew*

Signals and Systems Group,  
Department of Electronics and Electrical Engineering,  
University of Edinburgh, The King's Buildings,  
Edinburgh, EH9 3JL, Scotland, UK

Email : Charles.Unsworth@ee.ed.ac.uk  
Tel : (+44)-131-650-5596, Fax : (+44)-131-650-6554

*Keywords:* Radar, Chaos, Sea Clutter, k-distribution.

## ABSTRACT

There is current debate in the radar community whether sea clutter is stochastic or chaotic. In this paper, a stochastic k-distributed surrogate is generated for a typical sea clutter data set. The k-distributed set was then analysed using the methods recently applied to sea clutter by Haykin et al. The k-distributed set is shown to have DML and FNN values in the same range as reported by Haykin et al. and with positive and negative Lyapunov exponents. In addition, various white and correlated noise distributed sets are analysed in the same way and found to produce similar artefact. It is concluded that these chaotic invariants cannot be used to distinguish between chaotic and stochastic timeseries and are redundant in an application, such as radar sea clutter, where the time series is unknown and could be of a stochastic nature.

## 1. INTRODUCTION

There is current debate in the radar community whether sea clutter is stochastic or chaotic. Conventionally, high resolution radar sea clutter has been modelled by a stochastic compound k-distribution[1]. Recently, Haykin et al.[2][3] has performed a nonlinear analysis on sea clutter data sets and claims that sea clutter is a chaotic process. This nonlinear analysis hinges around two main chaotic invariants. These are the 'maximum likelihood estimation of the correlation dimension' (DML value)[4] and 'false nearest neighbours' (FNN)[5]. The Lyapunov exponents, are also measured, where the num-

ber of exponents to be measured is determined from the FNN calculation.

In [3] it was reported that sea clutter had fractional DML values in the range 4.1-4.5 and FNN global dimension in the range 5-6. It could be inferred from these results that the system is low dimensional and fractal which is symptomatic of chaos. From the FNN result reported in [3], 5-6 Lyapunov exponents were measured which gave positive and negative values, where one positive and negative exponent signifies chaos. In this paper we wish to test the robustness of the above mentioned chaotic invariants to stochastic time series and in particular a time series drawn from a k-distribution. If the chaotic invariants are robust they will be able to distinguish a stochastic time series from a chaotic one.

The paper is structured as follows.

- DML and FNN analysis for white stochastic time series;
- DML and FNN analysis for correlated Gaussian noise;
- Mutual information, autocorrelation, DML, FNN & Lyapunov exponents for a k-distributed surrogate.

## 2. WHITE NOISE SIGNALS

White noise signals are essentially high dimensional in the sense that high DML values and a high FNN global dimension are to be expected for such a series. Four white stochastic systems were generated. The signals generated were gamma, uniform, Gaussian and k-distributed. Each consisted of 50,000 data points which is equivalent to the length of the data record in [3]. The correlation dimension ( $D_2$ ) and 'maximum likelihood estimate of the correlation dimension' (DML

---

This work was supported by BAE Systems, DERA Malvern, EPSRC and the Royal Society. Sea clutter data was provided courtesy of DERA, Malvern.

value) which is a noise robust version of  $D_2$  were estimated using a method by Schouten et al.[4] which was employed in [3]. The results are shown in Table 1.

Data set	$D_2$	DML
Gamma (white)	1.38	1.38
Uniform (white)	4.20	4.20
Gaussian (white)	4.22	4.42
K (white)	1.94	1.94

Table 1

The results are somewhat alarming. In the case of the original method of Grassberger and Procaccia[6] the  $D_2$  value would become infinite for white noise since the noise fills high dimensional space resulting in a very large  $D_2$ . Schouten's method does not demonstrate this divergence, instead it suggests the data has a low fractal dimension which might be interpreted as the presence of chaos. Even more worrying is that for the uniform and Gaussian white noise signals the  $D_2$  and DML values are in the same range as in [3] which were measured for sea clutter. The FNN results for the parameter  $R_{TOL} = 10$ , as in [3], are shown in Figure 1. The algorithm designed by Kennel[7] was used to measure the FNN. An embedding delay of unity was used for the white stochastic sets.

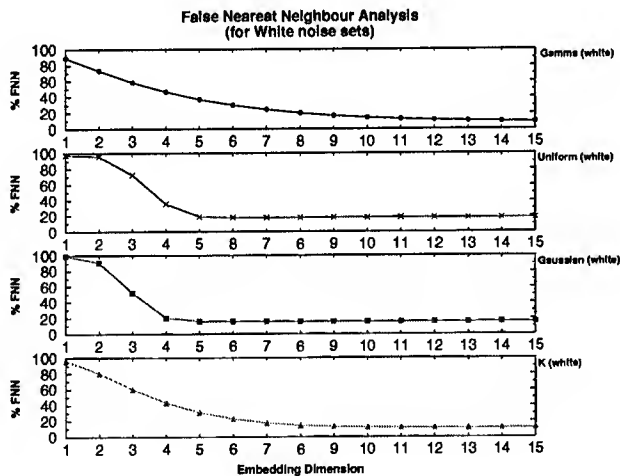


Figure 1:

The FNN for the gamma distribution would be expected result for a typical white noise system (i.e. monotonically decreasing). The white K distribution actually saturates at a global dimension of 11 which is maybe also acceptable for a noise system of 50,000 points. However, the FNN for the uniform and Gaussian white noise systems drop down to a saturation level at a global dimension of 5 which is the same as reported in [3] for sea clutter. From Abarbanel[5] this

result could be misconstrued as a low dimensional attractor with observational noise. Evidently the FNN's also generate artifact for white stochastic time series.

### 3. CORRELATED GAUSSIAN NOISE

What now follows is the same analysis applied to correlated stochastic time series. For this analysis, 50,000 point data records of correlated Gaussian noise for correlation coefficient  $\rho=0.1, 0.3, 0.5, 0.7, 0.9, 0.99$  and  $0.999$  were generated. The  $D_2$  and DML values together with the vector length(m) are shown in Table 2.

$\rho$	$D_2$	DML	Vector Length(m)
0.1	4.72	5.42	5
0.3	4.72	5.42	5
0.5	4.82	6.27	6
0.7	5.37	7.38	8
0.9	4.98	8.49	14
0.99	3.56	6.59	46
0.999	2.28	3.49	142

Table 2

As the correlation coefficient,  $\rho$ , increases so does the  $D_2$  and DML value. This continues to a ceiling at  $\rho=0.7-0.9$  and then decreases at high correlation values. The vector length(m) is also presented. Essentially from Schouten's work[4], random pairs of vectors are chosen of a particular vector length (m). And the maximum norm distance of 10,000 pairs of vectors is plotted as a cumulative histogram from which  $D_2$  and DML are estimated from. It can be seen that the vector length increases with the degree of correlation of the time series. It is crucial that m is estimated correctly. The vector length is inversely proportional to and derived from the number of crossings of the mean of the time series. (i.e. the larger the no. of crossings the smaller m will be). An assumption has been made here which is that there is some structure/repetition in the motion of the system (i.e. there exists orbits of an attractor). In white stochastic systems where there is no structure artifact must occur. A mean level will still exist and be crossed very frequently. However, these crossings are not structured, as in the orbits of an attractor, but are purely random. Therefore, a low m will be determined and a spurious estimate of  $D_2$  and DML will be made. As the correlation of the noise is increased more apparent structure appears since points in the time series become more dependent on the sample behind. Less crossings of the mean will occur and a larger m will be measured. This is evident in the Table 2. Therefore, for correlated noise signals the  $D_2$  and

DML values of [5] suggests low fractal dimensionality which could be mistaken as evidence for chaos. In order for Schouten's et al. method to work for stochastic systems the vector length ( $m$ ) must be calculated in another manner which is able to distinguish in some way if the system has true orbits occurring or not. The FNN results for the same correlated Gaussian time series are shown in Figure 2 for an  $R_{TOL} = 10$  and embedding delay of unity.

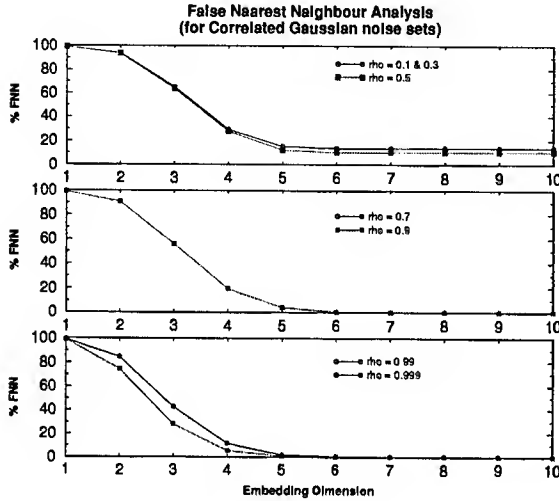


Figure 2:

At small correlation values of  $\rho$  the system resembles a low order 5 system with observational noise on top. As the correlation gets larger it appears as if the system has real dynamics. Therefore, the FNN seems to produce more artifact for heavily correlated stochastic time series.

#### 4. A K-DISTRIBUTED TIME SERIES

What follows is a time series analysis, using the techniques and parameters as described in [3], for a surrogate compound K-distributed time series.

Parameter	Description/value
Frequency	3GHz
Pulse compression	not used
Pulse width	1 $\mu$ s
Resolution	150m
Windspeeds	12.8m/s (VV)
Sea states	"strong breeze" (VV) (Beaufort scale: 6)
Polarisation/channel	VV (Agility not used)/ Q
PRF	20kHz
Grazing angle	0.12°
Beamwidth	6°

Table 3

The radar parameters of the actual sea clutter set are given in Table 3. The shape and scale parameters of the sea clutter were 23 and 661 respectively. A  $3 \times 10^6$  compound K-distributed surrogate data set was generated. This was achieved using Tough and Ward's method[8]. This technique enables the generation of surrogate data sets which are matched in both point probability density function and in autocorrelation sequence to the observed sea clutter. Figure 3. shows the time series for both the original sea clutter set and of the generated compound K-distributed set.

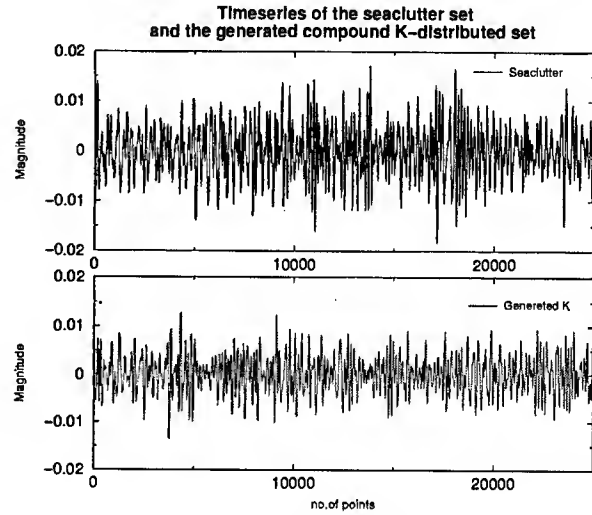


Figure 3:

The autocorrelation function(ACF) and mutual information (MI)[5] plots of the surrogate data were measured using 50,000 data points and are shown in Figure 4.

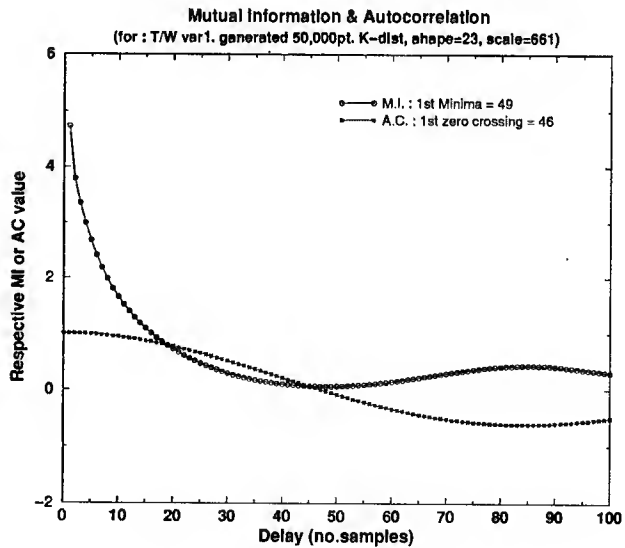


Figure 4:

The MI and ACF curves were found to be smooth. A zero crossing of the ACF=46 and 1st minima of the MI=49 were found to occur at roughly the same position. The  $D_2$  and DML values were measured for the 50,000 point set, the  $3 \times 10^6$  point set and also for 100 sets of 30,000 points in order to obtain a range. 10,000 vectors, specified by Schouten, were used for the 50,000 and 30,000 point data sets in order to estimate the  $D_2$  and DML values. This number of vectors used was appropriately scaled for the 3 million point set to 600,000 vectors. The results are shown in Table 4.

Data set	$D_2$	DML
50,000pts.	3.27	4.62
3 million pts.	3.78	4.55
100 sets of 30,000pts.	$3.11 \leq D_2 \leq 4.03$	$4.20 \leq DML \leq 5.1$

Table 4

Hence, the range of  $D_2$  and DML values are similar to the range that was reported for a number of clutter sets in [3].

The FNN results are shown in Figure 5 for both the  $R_{TOL} = 10$  and  $R_{TOL} = 25$  that is used in [3]. An embedding delay=MI=49 was chosen. The same results as reported in [3] were found (i.e. No noise floor is present and a global dimension of 5).

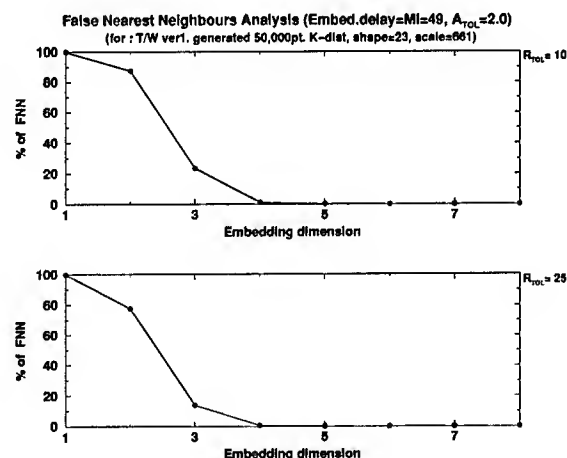


Figure 5:

Finally the Lyapunov exponents were measured using the result of the FNN calculation, as performed in [3], to determine the number of Lyapunov exponents to use. The algorithm by Ushaw[9] was used to measure the exponents. The algorithm is an extension of the Darbyshire and Broomhead[10] model. The Ushaw model reverts to the Darbyshire and Broomhead model when the 'number of B vectors in the average' is set to

unity. At higher values of this parameter local noise reduction takes place. Parameters for the Lyapunov calculation were then determined. They were found to be (svd window length=120, calculation period =1000, global embedding dimension=5, local embedding dimension=5, no.steps between re-initialisations=40, no of B vectors=60). Two calculations were made. The first with (number of B vectors in the average=1 i.e Darbyshire and Broomhead model) shown in Table 5.

<i>Lya.Exp.</i>	<i>nats/sample</i>
L1	+0.001112
L2	-0.003031
L3	-0.005450
L4	-0.010183
L5	-0.025041

Table 5

The second calculation was made with noise reduction (number of B vectors in the average=7) shown below in Table 6.

<i>Lya.Exp.</i>	<i>nats/sample</i>
L1	+0.002018
L2	-0.004457
L3	-0.005818
L4	-0.009299
L5	-0.025878

Table 6

For both calculations the Lyapunov exponents are very small, like those reported in [3], and positive and negative exponents were found that hallmark a chaotic system.

## 5. CONCLUSION

In conclusion, both white and correlated stochastic time series have been analysed using Schouten's correlation dimension estimate and the FNN test. In both cases, the tests resulted in the misclassification of a stochastic time series as a chaotic process. Low fractal  $D_2$  and DML values were generated from Schouten's method which are signatures of chaos. It is believed that this is due to the method of calculation of the vector length ( $m$ ). The FNN also provide low dimensional estimates which can be misinterpreted as an attractor. Compound k-distributed surrogate data was generated and passed through the same nonlinear analysis as was performed in [3] for sea clutter. Smooth Autocorrelation and Mutual information curves were measured. The  $D_2$  and DML, FNN values were found be similar to

those reported in [3]. Finally the Lyapunov exponents were measured and positive and negative exponents were found which suggest evidence of chaos. Clearly, the generated K-distributed data is not chaotic but the nonlinear tools used in [3] in the analysis of sea clutter suggest otherwise. This counter example demonstrates that it is not possible to distinguish a chaotic time series from a stochastic one using these invariants. Hence, it is suggested that these tools be used for deterministic time series only. Thus, we conclude that these chaotic invariants are redundant in an application, such as radar sea clutter, where the time series is unknown and could be of a stochastic nature. This reopens the question as to what the true nature of sea clutter actually is.

## REFERENCES

- [1] S.Watts, C.J.Baker, and D. K, "Maritime surveillance radar part 2: detection performance prediction in seaclutter," *IEE Proc. Part F.*, vol. 137, pp. 63-72, 1990.
- [2] S.Haykin and X.B.Li, "Detection of signals in chaos," *Proc.of IEEE*, vol. 83, pp. 95-122, 1995.
- [3] S.Haykin and S.Puthusserypady, "Chaotic dynamics of seaclutter: an experimental study," *Proc.of IEE Radar 1997*, pp. 75-79, 1997.
- [4] J.C.Schouten, F.Takens, and C. den Bleek, "Estimation of the dimension of a noisy attractor," *Phys.Rev.E*, vol. Vol.50, No.3, 1994.
- [5] H. Abarbanel, "Analysis of observed chaotic data," *Springer, New York*, 1996.
- [6] P.Grassberger and I.Procaccia, "Measuring the strangeness of strange attractors," *Physica 9D*, pp. 189-208, 1983.
- [7] H. A. M.B.Kennel, R.Brown, "Determining embedding dimensions for phase space reconstruction using the method of false nearest neighbours," *Dept.of physics, University of California, San Diego, Mail code R-002, La Jolla, CA 92093-0402*, 1992.
- [8] R.J.A.Tough and K.D.Ward, "The generation of correlated k-distributed noise," *Tech.Rep. DRA/CIS/CBC3/WP94001/2.0, Defence Research Agency, Farnborough, Hampshire, England. GU14 6TD*, 1994.
- [9] G.Ushaw, "Phd thesis - sigma delta modulation of a chaotic signal," *The University of Edinburgh, Edinburgh, Scotland, UK. EH9 3JL*, 1996.
- [10] A.G.Darbyshire and D.S.Broomhead, "The calculation of liapunov exponents from timeseries data," *Physica D89, no.3-4*, pp. 287-305, 1996.



# RECURSIVE ESTIMATOR FOR SEPARATION OF ARBITRARILY KURTOTIC SOURCES

*Mihai Enescu and Visa Koivunen*

Signal Processing Laboratory  
Helsinki Univ. of Technology  
P.O. Box 3000, FIN-02015 HUT, Finland

## ABSTRACT

Blind source separation has many important applications in communications and array signal processing. Many widely used methods require prior knowledge on the sign of the kurtosis of the sources and may fail if the mixtures contain both sub- and super-Gaussian signals. In this paper we present an adaptive algorithm for separating arbitrarily kurtotic sources. The blind separation problem is modeled using a state-space formulation. The resulting separation algorithm uses a subspace tracker and a predictor-corrector filter structure related to the well-known Kalman filter. It lends itself easily to real-time implementation. The zero-memory nonlinearities needed for finding independent sources are selected online by monitoring the statistics of each estimated source signal. Consequently, separation may be achieved even if a change in the sign of the kurtosis occurs. Simulation examples illustrating the ability to adapt to time-varying mixing systems and source distributions of unknown kurtosis are presented using communications and biomedical signals.

## 1. INTRODUCTION

Blind Source Separation (BSS) has important applications in biomedical signal analysis, communications and array signal processing. Adaptive separation methods are often required because mixing system and signal or noise statistics may be time-varying. Moreover, real-time computation is desirable in many key application areas.

Typically blind separation algorithms based on higher order statistics assume that the sign of the kurtosis is known and the same for all sources. Consequently, the zero-memory nonlinearity employed is fixed in advance. The choice of nonlinearity is critical for achieving the separation. One way to overcome this problem is to

approximate the nonlinear function by a linear combination of sigmoids with adjustable slope and bias, thus avoiding the estimation of the kurtosis [10]. A large number of parameters have to be adapted which leads to high computational complexity. Moreover, deriving recursive algorithms needed in real-time operation may be tedious.

In this paper, we propose an algorithm for online separation of source signals with arbitrary kurtosis. Separation is performed by employing a subspace tracker and a recursive estimator with a predictor-corrector form. Instead of making restrictive assumptions on source pdf's and consequently fixing the nonlinearity in advance, the output statistics of each channel is monitored and the nonlinearity is selected appropriately. Consequently, a fully adaptive algorithm for blind separation is obtained that easily lends itself to real-time implementation. The performance of the proposed method is studied in simulations where sources with different kurtosis parameters are used and the mixing system is time-varying.

This paper is organized as follows. The BSS problem is presented first. Then there is a brief description of the recursive estimator for blind separation and discussion of selecting appropriate nonlinearities on-line. In section 4, examples on separating mixtures of both sub- and super-Gaussian sources are given.

## 2. BLIND SEPARATION

Over the last few years BSS has received a lot of attention in the signal processing, communications and neural network research communities (see [2],[3] and references therein). The observed noisy mixtures and the unobserved source signals are related by

$$\mathbf{z}(k) = \mathbf{A} \mathbf{s}(k) + \mathbf{v}(k) \quad (1)$$

where  $\mathbf{A}$  is an  $n \times m$  matrix of unknown mixing coefficients,  $n \geq m$ ,  $\mathbf{s}$  is a column vector of  $m$  source signals,

---

This work was funded by the Academy of Finland

$\mathbf{z}$  is a column vector of  $n$  mixtures,  $\mathbf{v}$  is an additive noise vector and  $k$  is the time index. The mixing is assumed to be instantaneous and matrix  $A$  is assumed to be of full rank. Source signals are typically assumed zero mean and stationary.

The separation task at hand is to estimate a separating matrix  $W$  or mixing matrix  $H$  so that the original sources are recovered from the noisy mixtures. Prior to separation, the observed signals are typically spatially whitened and the signal powers are normalized to unity. By projecting the input data  $\mathbf{z}$  into an  $m$ -dimensional signal subspace yielding  $\mathbf{y}$ , the problem becomes easier to solve because  $n = m$  and  $m \times m$  separating matrix will be orthogonal, i.e.,  $W = H^{-1} = H^T$ . Moreover, some noise is also attenuated. An estimate  $\mathbf{x}$  of unknown sources  $\mathbf{s}$  may then given by

$$\hat{\mathbf{s}} = \mathbf{x} = \hat{H}^T \mathbf{y}. \quad (2)$$

where  $\mathbf{y}$  is the whitened data. The estimate can be obtained only up to a permutation and scaling of  $\mathbf{s}$ .

### 3. ADAPTIVE SEPARATION OF ARBITRARILY KURTOTIC SOURCES

The goal of the adaptive blind algorithm presented in this section is to separate sources with arbitrary kurtosis parameters. In this algorithm a zero-memory non-linearity is used. The type of nonlinearity is selected adaptively based on the statistics of the output. The actual adaptive algorithm consists of two parts: a signal subspace tracker and a recursive predictor-corrector filter structure. This type of structure allows for real-time implementation.

#### 3.1. Signal Subspace Tracking

In order to make the separation problem easier, adaptive signal subspace tracking is employed. The  $n$ -dimensional observations  $\mathbf{z}(k)$  are projected along eigenvectors corresponding to  $m$  largest eigenvalues. Signal subspace eigenvectors  $U$  and eigenvalues  $\Lambda$  are tracked on-line using the adaptive algorithm introduced in [6]. Estimates of the signal subspace eigenvectors and noise variance are updated at the arrival of each new observation vector. Thus, at each step  $k$  we obtain a whitened data vector  $\mathbf{y}(k)$  by applying the transformation  $R(k) = \Lambda^{-1/2} U^T$  to the observation vector  $\mathbf{z}(k)$ . In case abrupt change in the mixture covariance structure occurs, the subspace tracker is reinitialized so that recent observations are trusted more [8].

#### 3.2. Separation algorithm

The actual separation algorithm can be considered a modified Kalman filter presented in a predictor-corrector form. This is achieved by describing the blind source separation problem using a state-space model:

$$\mathbf{x}(k) = F(k|k-1)\mathbf{x}(k-1) + G(k)\mathbf{w}(k-1) \quad (3)$$

$$\mathbf{y}(k) = H(k)\mathbf{x}(k) + \mathbf{v}(k) \quad (4)$$

where  $\mathbf{x}$  is the state vector to be estimated and  $\mathbf{y}$  is the whitened observation vector. The noise sequences  $\mathbf{w}$  and  $\mathbf{v}$  are Gaussian white, mutually uncorrelated with covariance matrices  $Q(k)$  and  $R(k)$ . The measurement noise variance is estimated using the subspace tracking algorithm. Having the state-space model, the predicted state estimate  $\hat{\mathbf{x}}(k|k-1)$  is given by:

$$\hat{\mathbf{x}}(k|k-1) = F(k|k-1)\hat{\mathbf{x}}(k-1|k-1) \quad (5)$$

The correction equations update the predicted state estimate to a filtered state estimate based on the new information conveyed by the measurements. The estimated source signals are given by:

$$\hat{\mathbf{x}}(k|k) = \hat{\mathbf{x}}(k|k-1) + K(k)[\mathbf{y}(k) - H\hat{\mathbf{x}}(k|k-1)] \quad (6)$$

where  $K(k)$  is the Kalman gain. The prediction and correction error covariance matrices  $P(k|k-1)$  and  $P(k|k)$  are updated as well (see [5]).

In the BSS problem, matrices  $H$  and  $F$  are not known and have to be estimated simultaneously. Matrix  $H$  describes the mixing system whereas matrix  $F$  models how sources evolve over time. Structure of the  $F$  matrix depends on the application. For example, in some cases the source signals may exhibit an autoregressive structure. In this case, in order to make the prediction more accurate, matrix  $F$  may be augmented to contain a low order AR model. This also allows for noise attenuation. The update equations for the prediction and correction error covariance matrices are changed [8].

In estimating the mixing matrix  $H$ , separate expressions for innovation and gain for correcting the elements of  $H$  are required. The "innovation" in estimating  $H$  is

$$\tilde{\mathbf{y}}_H(k) = \mathbf{y}(k) - H(k-1)\hat{\mathbf{x}}_H(k) \quad (7)$$

where  $\hat{\mathbf{x}}_H(k) = \mathbf{g}(\mathbf{u}(k))$ ,  $\mathbf{u}(k) = H^T \mathbf{y}(k)$ ,  $\mathbf{g}(\mathbf{u}(k)) = [g_1(u_1(k)) \dots g_m(u_m(k))]^T$ , and  $g_i(\cdot)$  are nonlinear contrast functions. In the subspace tracking stage, the components of  $\mathbf{y}$  are normalized to have unit variance. The gain  $K_H$  used in estimating the mixing matrix is

$$K_H = \frac{P\hat{\mathbf{x}}_H}{\hat{\mathbf{x}}_H^T P \hat{\mathbf{x}}_H + 1}. \quad (8)$$

Finally, the update for  $H$  is given by:

$$H(k) = H(k-1) + (y - H(k-1)\hat{x}_H)K_H^T \quad (9)$$

The adaptation rate for  $H$  should be relatively slow compared to the adaptation rate for the actual state variables to have better stability. At each time step  $k$  we have an estimate of the source signals and of the mixing matrix.

### 3.3. Selecting appropriate nonlinearity

Typically prior information on the sign of the kurtosis is assumed to be available and the nonlinearities  $g_i(\cdot)$  are selected accordingly. This assumption is often unreasonable. In order to effect a truly blind algorithm, the statistics of each output of the separation system are recursively tracked and an appropriate nonlinearity for each channel is selected from two alternatives depending on whether the source is deemed to have negative or positive kurtosis. The selection principle employed here was introduced in [4] and stems from the stability analysis presented in [1]. In order to choose the nonlinear functions for each channel, at each time step  $k$  we recursively estimate the following statistics:

$$\begin{aligned} \sigma_i^2(k) &= (1 - \mu)\sigma_i^2(k-1) + \mu|\hat{x}_i(k)|^2 \quad (10) \\ \kappa_{i,r}(k) &= (1 - \mu)\kappa_{i,r}(k-1) + \mu g_r'(\hat{x}_i(k)) \\ \rho_{i,r}(k) &= (1 - \mu)\rho_{i,r}(k-1) + \mu\hat{x}_i(k)g_r(\hat{x}_i(k)) \end{aligned}$$

where  $1 \leq i \leq m$ ,  $r = \{1, 2\}$  refers to the type of nonlinearity and  $\mu$  is a positive constant such that  $0 < \mu \ll 1$ .

$\sigma_i^2(k)$ ,  $\kappa_{i,r}(k)$  and  $\rho_{i,r}(k)$  can be defined as:

$$\sigma_i^2(k) = E\{\hat{x}_i^2(k)\} \quad (11)$$

$$\kappa_{i,r}(k) = E\{g_r'(\hat{x}_i(k))\} \quad (12)$$

$$\rho_{i,r}(k) = E\{\hat{x}_i(k)g_r(\hat{x}_i(k))\} \quad (13)$$

where  $\hat{x}_i(k)$  is the source signal estimate on channel  $i$  at time  $k$ . Let us denote  $\mathcal{K}_1 = \sigma_i^2(k)\kappa_{i1}(k) - \rho_{i1}(k)$  and  $\mathcal{K}_2 = \sigma_i^2(k)\kappa_{i2}(k) - \rho_{i2}(k)$ . The nonlinear function for the  $i$ th component at time  $k$  is selected as follows:

$$g_{ik}(x) = \begin{cases} g_1(x), & \text{if } \mathcal{K}_1 - \mathcal{K}_2 < 0 \\ g_2(x), & \text{otherwise} \end{cases} \quad (14)$$

The functions  $g_1$  and  $g_2$  are the corresponding nonlinearities if the sources are sub- or super-Gaussian. There are many different contrast functions that may be used in order to perform the separation (see [3]). It has been proven [7] that for the nonlinear PCA class of algorithms suitable nonlinearities are odd polynomial

functions in the case of positive kurtotic sources and hyperbolic tangents in the case of negative kurtotic. In this paper  $g_{i,1}(u) = \tanh(\alpha_1 u)$  and  $g_{i,2}(u) = \alpha_2 u^3$  are employed, where  $\alpha_1$  and  $\alpha_2$  are constants. Thus, at each time  $k$ ,  $g_i$  component from  $\mathbf{g}(\mathbf{u}(k))$  is either  $g_{i,1}(\cdot)$  or  $g_{i,2}(\cdot)$  based on criterion (14).

## 4. EXAMPLES

In this section, the separation performance of the proposed BSS algorithm is studied in simulations. In order to demonstrate the practical applicability of the algorithm we use ECG signals. In this example, 2000 samples of  $m=3$  source signals and  $n=4$  mixtures are used. The initial sources are two positive kurtotic ECG signals representing maternal and fetal heart beats at frequencies slightly above 1 Hz and slightly below 3 Hz respectively and a interfering sinusoid of 50 Hz (negative kurtotic). In practice this problem may be encountered if the electrocardiograph is disturbed by some unwanted interference due to poor grounding or muscle contraction during the measurements. The mixing coefficient matrix  $A$  is randomly generated and the observed mixtures are contaminated with zero mean additive Gaussian noise with  $\sigma = 0.1$ . A low order ( $p=2$ ) AR model is employed in the state prediction matrix  $F$ . The constants used in the contrast functions are  $\alpha_1 = 1$  and  $\alpha_2 = 1/3$ . The type of nonlinearity needed in separation is selected on-line using the criterion given in (14). At each time step  $k$  we update the statistics per (10), with  $\mu = 0.01$ . The results of the separation are presented in Fig. 1. In order to qualitatively illustrate the recovery of the shape of the ECG signals with high fidelity, original noise free sources and separated sources are plotted in Fig. 2.

The selection of the appropriate zero-memory nonlinearity is simulated next. The track of the sign of  $\mathcal{K}_1 - \mathcal{K}_2$  for each channel is presented in Fig. 3. If  $\mathcal{K}_1 - \mathcal{K}_2$  is positive it means that on the respective channel we have a super-Gaussian signal, otherwise we have a sub-Gaussian signal. Typically number of samples needed to achieve separation is 300.

A non-stationary scenario may be simulated as follows: given the stationary input source signals we perform the mixing by using a slowly time-varying mixing matrix. This can be obtained by applying a rotation matrix  $T(\theta(k))$  to the random mixing matrix  $A$ , where  $T(\theta(k))$  is given by:

$$T(\theta(k)) = \begin{bmatrix} \cos(\theta(k)) & \sin(\theta(k)) & 0 & 0 \\ -\sin(\theta(k)) & \cos(\theta(k)) & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \quad (15)$$

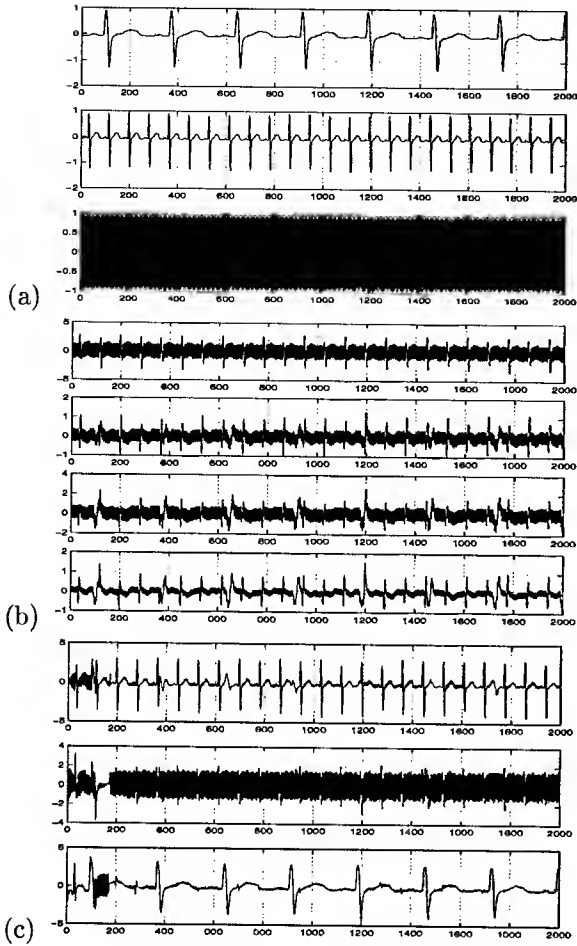


Figure 1: An example of blind separation from noisy mixtures. (a) Noise free source signals, (b) noisy mixtures with additive Gaussian noise with zero mean and  $\sigma = 0.1$ , (c) separation results obtained using on-line eigenpair tracking and predictor-corrector structure by adaptively selecting the appropriately nonlinearity for each channel.

The subspace tracking algorithm requires that the number of sensors is greater than the number of sources ( $n > m$ ). In this example the same ECG and sinusoidal  $m=3$  source signals of 2000 samples and  $n=4$  mixtures are used. To provide an initial estimate of the mixing matrix and predictor-corrector parameters, static random mixing matrix  $A$  is used for the first 1000 samples. For the next 1000 samples the angle  $\theta(k)$  is linearly changed from the initial value  $\theta(1000) = 0$  to the final value  $\theta(2000) = \pi/3$ . The separation result is presented in Fig. 4.

Good estimates of the source signals are possible due to the ability of the subspace algorithm to track the eigenvalues and eigenvectors in non-stationary environment. The convergence of the tracking method

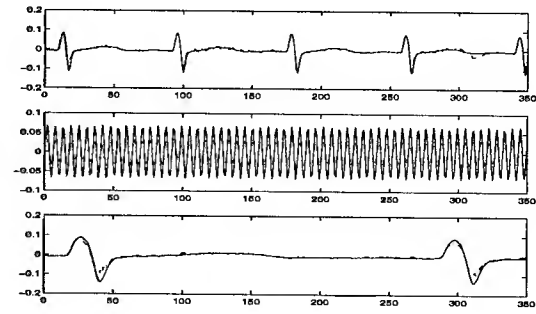


Figure 2: Recovering the original source with high fidelity: Points 1430 - 1780 from the original signals (continuous line) and the recovered sources (dash-dot line)

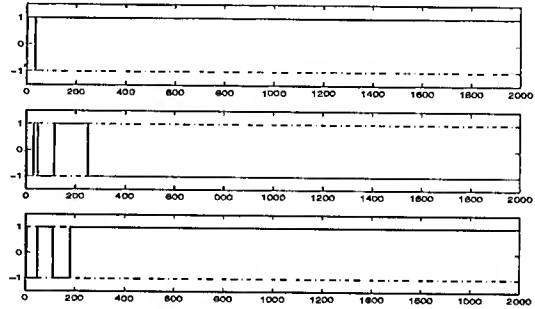


Figure 3: Trace of the decision criterion  $\mathcal{K}_1 - \mathcal{K}_2$  for each channel. The positive value stands for super-Gaussian, the negative for sub-Gaussian.

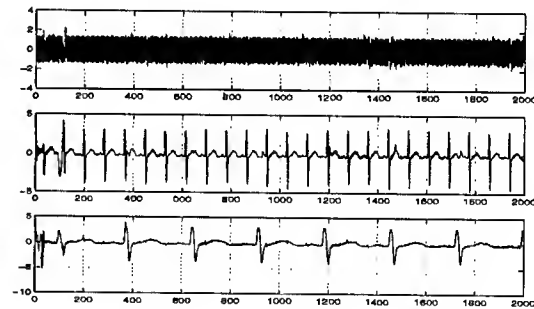


Figure 4: BSS for a slowly time-varying mixing structure.

is illustrated by plotting the canonical angles between the basis vectors in the estimated and theoretical signal subspaces. We consider  $\hat{\mathbf{u}}$  and  $\mathbf{u}$  being the eigenvectors of the estimated and true signal subspaces and we compute the singular value decomposition of  $\hat{\mathbf{u}}^T \mathbf{u}$ . Let  $\gamma_1 > \gamma_2 > \dots > \gamma_m$  be the singular values of  $\hat{\mathbf{u}}^T \mathbf{u}$ . The canonical angles between the basis vectors are obtained by  $\angle(\hat{\mathbf{u}}, \mathbf{u}) = \cos^{-1} \gamma_i$ . If the maximum canonical angle is small the subspaces are close to each

other. The maximum canonical angle for the mixtures used in Fig. 1 is shown in Fig. 5. The results indicate that the subspace tracker converge relatively fast to true signal subspace.

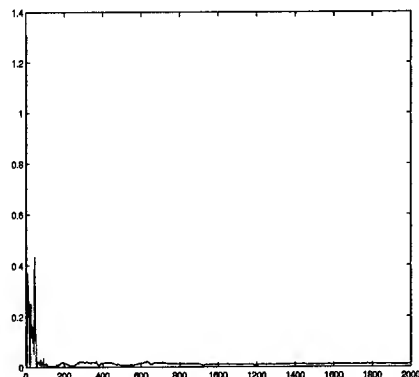


Figure 5: Maximum canonical angle between estimated and theoretical signal subspace basis vectors.

The performance of the proposed method is also investigated using communications signals. 4-QAM, 16-QAM (sub-Gaussian) and Laplacian distributed jamming signals (super-Gaussian) are randomly mixed and contaminated with Gaussian noise (SNR = 32dB). The Laplacian p.d.f. is given by  $p(s) = 0.5e^{-|s|}$ . The total number of samples is 2000 and the number of receivers is four and is not changed during the simulation. In this case the state prediction matrix is  $F = I$ . From the signal space diagram of the mixed signals no QAM constellation can be distinguished. The result of the separation is presented in Fig. 6.

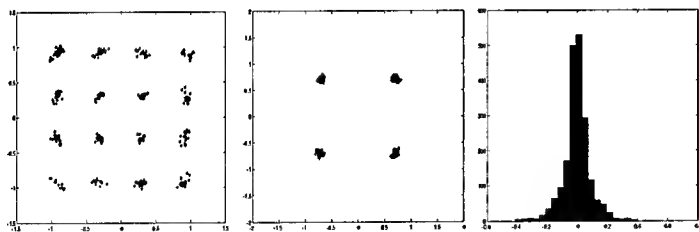


Figure 6: Separation result in the case of communications signals. Only the last 300 samples are shown.

## 5. CONCLUSION

In many BSS problems we do not have prior information on the type of pdf and the sign of the kurtosis. Furthermore, signal statistics may be time varying. We introduced an algorithm that does not make restrictive assumptions on the form of the pdf, can adapt to changes in the mixing system and signal statistics, and

lends itself to real-time computation. The algorithm performs signal subspace tracking and employs a recursive estimator to produce estimates of the source signals at the arrival of each new mixture observation. A zero memory nonlinearity is employed in separation. The type of nonlinearity is adaptively determined based on the statistics computed from the output.

## REFERENCES

- [1] S-I. Amari, T.-P. Chen, A. Cichocki, "Stability analysis of adaptive blind source separation", *Neural Networks*, Vol 10, No. 8, pp. 1345-1351, 1997.
- [2] S-I. Amari, A. Cichocki, "Adaptive Blind Signal Processing-Neural Network Approaches", *Proceedings of the IEEE*, Vol 86, No. 10, pp. 2026-2048, 1998.
- [3] J-F. Cardoso, "Blind Signal Separation: Statistical Principles", *Proceedings of the IEEE*, Vol 86, No. 10, pp. 2009-2025, 1998.
- [4] S.C. Douglas, A. Cichocki, S-I. Amari, "Multichannel blind separation and deconvolution of sources with arbitrary distributions", *Neural Networks for Signal Processing, Proc. IEEE Workshop (NNSP-97)*, pp. 436-445, 1997.
- [5] S. Haykin, "Adaptive Filter Theory", 3rd edition, Prentice-Hall, 1996.
- [6] I. Karasalo, "Estimating the Covariance Matrix by Signal Subspace Averaging", *IEEE T. Acoustics, Speech, and Signal Processing (T-ASSP)*, Vol. 34, No. 1, pp. 8-12, 1986.
- [7] J. Karhunen, E. Oja, L. Wang, R. Vigario, J. Joutsensalo, "A Class of Neural Networks for Independent Component Analysis", *IEEE Trans. on Neural Networks*, Vol. 8, No. 3, pp. 486-504, 1997.
- [8] V. Koivunen, M. Enescu, E. Oja, "Adaptive Algorithm for Blind Separation from Noisy Time-Varying Mixtures", *Neural Computation*, Submitted Nov. 1999.
- [9] M. Wax, T. Kailath, "Detection of Signals by Information Theoretic Criteria", *IEEE Trans. Acoustics, Speech, and Signal Processing (T-ASSP)*, Vol. 33, No. 2, pp. 387-392, 1985.
- [10] L. Xu, C.C. Cheung, S. Amari, "Learned Parametric Mixture Based ICA Algorithm", *Neurocomputing*, Vol. 22, pp. 69-80, 1998.

# A SECOND ORDER MULTI OUTPUT DECONVOLUTION (SOMOD) TECHNIQUE

*Hicham Bousbia-Salah and Adel Belouchrani*

Electrical Engineering Department,  
Ecole Nationale Polytechnique,  
P.O. Box 182 El-Harrach, Algiers 16200, Algeria.  
(e-mail: hichambs@yahoo.fr, belouchrani@hotmail.com)

## ABSTRACT

In this paper, we present an efficient solution to the blind multi-channel deconvolution problem that consists of recovering independent source signals from their convolutive mixtures. In the case of instantaneous mixtures, a robust solution referred to as Second Order Blind Identification (SOBI) has been proposed previously. It is based on the joint diagonalization of spatio-temporal correlation matrices. Herein, we extend this technique to the convolutive mixture case. In contrast to existing deconvolution techniques, this new approach is able to deal with an overestimated source number. The proposed method has been successfully applied to the deconvolution of speech signals.

## 1. INTRODUCTION

If we consider a set of received signals that are linear convolutive mixtures of decorrelated source signals, the objective of blind deconvolution is to recover the source signals from the set of received signals without any knowledge of the linear mixtures or the Linear Time Invariant (LTI) systems. For instantaneous mixtures, a Second Order Blind Identification (SOBI) algorithm has been presented [1] and showed to be very robust for temporally correlated sources. There are two ways to achieve blind deconvolution. One way is to first identify the channel system from the output mixtures and then to design an equalizer accordingly [2]. The other way consists of directly designing an equalizer from the output mixtures. This approach bypasses the problem of blind system identification and is less costly in computation. Using the second approach, we extend the SOBI technique to the convolutive mixture case. It is based on the joint diagonalization of spatio-temporal corre-

lation matrices. The proposed method has been successfully applied to the deconvolution of speech signals and showed to be robust with respect to additive noise. Furthermore, this new approach is able to deal with an overestimated source number. In the next section, we will present the data model, the different hypothesis and the identifiability conditions. The proposed algorithm will be described in section 3. And finally, some simulation results are provided in section 4.

## 2. PROBLEM FORMULATION

### 2.1. Data Model

Consider a discrete time multiple input multiple output (MIMO) linear time invariant model given by,

$$x_i(n) = \sum_{j=1}^M \sum_{l=0}^{L-1} h_{ij}(l) s_j(n-l) + n_i(n), \text{ for } i = 1, \dots, N \quad (1)$$

where  $s_j(n)$ ,  $j = 1, \dots, M$  are the  $M$  source signals (model inputs),  $x_i(n)$ ,  $i = 1, \dots, N$ , are the  $N$  sensor signals (model outputs) with  $N \geq M$ ,  $h_{ij}$  is the transfer function between the  $j$ -th source and the  $i$ -th sensor with an overall extend  $L$ , and  $n_i(n)$ ,  $i = 1, \dots, N$ , are additive white noises.

The assumptions made about the data model are as follows:

**A1)** The source signals  $s_j(n)$ ,  $j = 1, \dots, M$ , are mutually decorrelated and each source signal is temporally coherent.

**A2)** The noise processes  $n_i(n)$ ,  $i = 1, \dots, N$ , are zero-mean stationary processes independent of the source signals.

The purpose of blind multi channel deconvolution is to recover the source signals based only on the sensor signals. This leads to find a set of weights  $\{w_{ji}(l)\}$  such

---

The Authors would like to thank STEP Alger, distributor of Motorola in Algeria, for its support for the presentation of this work.

that,

$$\tilde{s}_j(n) = \sum_{i=1}^N \sum_{l=0}^{L-1} w_{ji}(l) x_i(n-l), \quad \text{for } j = 1, \dots, M \quad (2)$$

where  $\tilde{s}_j(n)$  are the recovered source signals.

We can rewrite equation (2) in the following matrix form,

$$\tilde{\mathbf{s}}(n) = \mathbf{W}\mathbf{x}(n) \quad (3)$$

where

$$\begin{aligned} \tilde{\mathbf{s}}(n) &= [\tilde{s}_1(n), \dots, \tilde{s}_M(n)]^T \\ \mathbf{x}(n) &= [x_1(n), \dots, x_1(n-L+1), \dots, x_N(n-L+1)]^T \\ \mathbf{W} &= \begin{bmatrix} w_{11}(0) & \dots & w_{11}(L-1) & \dots & w_{1N}(L-1) \\ w_{21}(0) & \dots & w_{21}(L-1) & \dots & w_{2N}(L-1) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ w_{M1}(0) & \dots & w_{M1}(L-1) & \dots & w_{MN}(L-1) \end{bmatrix} \end{aligned} \quad (4)$$

## 2.2. Identifiability

Using the z-transform, the design of an equalizer  $\mathbf{W}(z)$  that recovers the original source signals only from the observations  $\mathbf{x}(n)$  can be formulated as follows:

$$\tilde{\mathbf{s}}(n) = \mathbf{W}(z)[\mathbf{H}(z)\mathbf{s}(n) + \mathbf{n}(n)] \quad (5)$$

$$\tilde{\mathbf{s}}(n) = \mathbf{W}(z)\mathbf{H}(z)\mathbf{s}(n) + \mathbf{W}(z)\mathbf{n}(n) \quad (6)$$

Let us write,

$$\mathbf{G}(z) = \mathbf{W}(z)\mathbf{H}(z) \quad (7)$$

As shown in [3], the LTI system represented by its transfer function matrix  $\mathbf{G}(z)$  is said to be transparent or *decoupled* if  $\mathbf{G}(z)$  has a single nonzero monomial entry in each row and each column.

In other words, an LTI system is transparent if and only if  $\mathbf{G}(z)$  can be decomposed into:

$$\mathbf{G}(z) = \mathbf{\Delta}(z)\mathbf{D}\mathbf{P} \quad (8)$$

where  $\mathbf{\Delta}(z)$  is a diagonal matrix with diagonal entries:

$$\lambda_{ii} = z^{l_i} \quad (9)$$

where  $l_i$  is a non-negative integer,  $\mathbf{D}$  is a constant diagonal matrix, and  $\mathbf{P}$  a permutation matrix.

Then, a channel system  $\mathbf{H}(z)$  is said to be *deconvolvable* if there exists an equalizer  $\mathbf{W}(z)$  so that the composite system  $\mathbf{G}(z)$  is transparent.

Furthermore, a necessary and sufficient condition for  $\mathbf{H}(z)$  to be *deconvolvable* is that the greatest common divisor of all the minors of order  $M$  in  $\mathbf{H}(z)$  is nonzero monomial (see [4] for details).

## 3. THE PROPOSED ALGORITHM (SOMOD)

The problem of blind multi channel deconvolution is to find  $\mathbf{W}$  an  $[M \times NL]$  matrix such that  $\tilde{\mathbf{s}}(n) = \mathbf{s}(n)$ . We can define the source correlation matrices at time lag  $k$  as:

$$\mathbf{R}_s(k) = E[\mathbf{s}(n)\mathbf{s}(n-k)^*] \quad (10)$$

Where  $*$  denotes the transpose conjugate of a vector. Under relation (3), the above equation can be put in the following form:

$$\mathbf{R}_s(k) = \mathbf{W}\mathbf{R}_x(k)\mathbf{W}^H \quad (11)$$

where,

$$\mathbf{R}_x(k) = E[\mathbf{x}(n)\mathbf{x}(n-k)^*] \quad (12)$$

are the data correlation matrices at time lag  $k$ .

Let us consider the following decomposition of  $\mathbf{W}$ ,

$$\mathbf{W} = \mathbf{U}^H\mathbf{B} \quad (13)$$

where  $\mathbf{U}$  is an  $[M \times M]$  unitary matrix,  $^H$  denotes the transpose conjugate of a matrix and  $\mathbf{B}$  is an  $[M \times NL]$  matrix.

Substituting (13) into (11) and assuming, without loss of the generality, that the source signals are of unit variance<sup>1</sup>, one can write

$$\mathbf{R}_s(0) = \mathbf{B}\mathbf{R}_x(0)\mathbf{B}^H = \mathbf{I} \quad (14)$$

According to equation (14),  $\mathbf{B}$  is nothing than a whitening matrix that can be obtained from an eigen decomposition of  $\mathbf{R}_x(0)$ .

For time lag  $k$ ,  $k \neq 0$ , we have

$$\mathbf{R}_s(k) = \mathbf{U}^H\mathbf{B}\mathbf{R}_x(k)\mathbf{B}^H\mathbf{U} = \mathbf{\Delta}_k \quad (15)$$

where  $\mathbf{\Delta}_k$  is a diagonal matrix according to assumption A1). By denoting

$$\underline{\mathbf{R}}_x(k) = \mathbf{B}\mathbf{R}_x(k)\mathbf{B}^H \quad (16)$$

where  $\{\underline{\mathbf{R}}_x(k), k = 1, \dots, K\}$  is a set of  $K$  whitened data correlation matrices at different time lags, we obtain the following key relation

$$\mathbf{\Delta}_k = \mathbf{U}^H\underline{\mathbf{R}}_x(k)\mathbf{U} \quad (17)$$

Since the matrix  $\mathbf{U}$  is unitary and  $\mathbf{\Delta}_k$  is diagonal, expression (17) shows that any whitened data correlation matrix is diagonal in the basis of the columns of the matrix  $\mathbf{U}$  (the eigenvalues of  $\underline{\mathbf{R}}_x(k)$  being the diagonal entries of  $\mathbf{\Delta}_k$ ).

<sup>1</sup>Because of the well known ambiguity of blind identification.

If, for the time lag  $k$ , the diagonal elements of  $\Delta_k$  are all distinct, the missing unitary matrix  $\mathbf{U}$  may be 'uniquely' (i.e. up to permutation and phase shifts) retrieved by computing the eigen decomposition of  $\underline{R}_x(k)$ .

Indeterminacy occurs in the case of degenerate eigenvalues. It does not seem possible to *a priori* determine some value for the delay  $k$  such that the diagonal entries of  $\Delta_k$  are all distinct. Of course, if the source signals have different spectral shapes, such a kind of eigenvalue degeneracy is unlikely, but it is to be expected that when some eigenvalues of  $\underline{R}_x(k)$  comes close to degeneracy, the robustness of determining  $\mathbf{U}$  from eigen decomposition of a single whitened data correlation matrix is seriously impaired.

The situation is more favorable when considering simultaneous diagonalization of a set  $\{\underline{R}_x(k)\}$  of  $K$  whitened data correlation matrices. This set is simultaneously diagonalizable by the unitary matrix  $\mathbf{U}$  as in (17).

The matrix  $\mathbf{U}$  is unique (to a permutation matrix and phase factors) if, and only if, for any pair  $(i, j)$  of sources, there exists a time lag  $k$  such that  $E[s_i(n)s_i(n-k)^*] \neq E[s_j(n)s_j(n-k)^*]$ . Of course, the simultaneous diagonalization holds only for the exact statistics; empirical statistics may only be *approximately* simultaneously diagonalized under the same unitary transform. This calls for the definition of the *approximate* simultaneous diagonalization.

**Joint diagonalization:** The joint diagonalization (JD) [1] can be explained by first noting that the problem of the diagonalization of a single  $n \times n$  normal matrix  $\mathbf{M}$  is equivalent to the maximization of the criterion [8]

$$C(\mathbf{M}, \mathbf{V}) \stackrel{\text{def}}{=} \sum_i |\mathbf{v}_i^* \mathbf{M} \mathbf{v}_i|^2 \quad (18)$$

over the set of unitary matrices  $\mathbf{V} = [\mathbf{v}_1, \dots, \mathbf{v}_n]$ . Hence, the joint diagonalization of a set  $\{\mathbf{M}_k | k = 1..K\}$  of  $K$  arbitrary  $n \times n$  matrices is defined as the maximization of the following JD criterion:

$$C(\mathbf{V}) \stackrel{\text{def}}{=} \sum_k C(\mathbf{M}_k, \mathbf{V}) = \sum_{k,i} |\mathbf{v}_i^* \mathbf{M}_k \mathbf{v}_i|^2 \quad (19)$$

under the same unitary constraint. An efficient joint approximate diagonalization algorithm exists in [1] and it is a generalization of the Jacobi technique [8] for the exact diagonalization of a single normal matrix.

Finally, the unitary matrix  $\mathbf{U}$  in (13) is obtained by the joint diagonalization of the set  $\{\underline{R}_x(k)\}$  which

corresponds to the maximization:

$$\hat{\mathbf{U}} = \underset{\mathbf{U}}{\text{Argmax}} \sum_{k=0}^{L-1} \sum_{i=1}^M |\mathbf{u}_i^* \underline{R}_x(k) \mathbf{u}_i|^2 \quad (20)$$

with  $\mathbf{U} = [\mathbf{u}_1, \dots, \mathbf{u}_M]$ .

## Implementation issue

The eigen decomposition of  $\mathbf{R}_x(0)$  for the determination of matrix  $\mathbf{B}$  will provide us with  $M \times L$  eigen vectors that span the extended subspace. Only  $M$  of them span the original source subspace. Hence, it is impossible, without any knowledge of the original sources to select the  $M$  eigen vectors among the  $M \times L$  obtained. We propose to use all the  $M \times L$  vectors in order to determine  $\mathbf{B}$  which will change the dimension of  $\mathbf{B}$  from  $[M \times NL]$  to  $[ML \times NL]$ . Then, we maximize the JD criterion using a unitary matrix  $\mathbf{U}'$  of dimension  $ML \times ML$ :

$$\hat{\mathbf{U}}' = \underset{\mathbf{U}'}{\text{Argmax}} \sum_{k=0}^{L-1} \sum_{i=1}^{ML} |\mathbf{u}'_i^* \underline{R}_x(k) \mathbf{u}'_i|^2 \quad (21)$$

with  $\mathbf{U}' = [\mathbf{u}'_1, \dots, \mathbf{u}'_{ML}]$ .

One can easily show that the maximization (21) leads to the maximization (20) and among the  $M \times L$  eigen vectors of  $\mathbf{U}'$ ,  $M$  of them correspond to the desired sources. The desired  $M$  eigen vectors are selected from the  $M \times L$  ones by choosing those which lead to the smallest correlation coefficients of the recovered signals. The proposed approach has shown to be robust with respect to an overestimated source number.

## 4. SIMULATIONS

### Example 1

we consider an array of 2 sensors receiving signals from 2 sources in the presence of white Gaussian noise. The channel length is  $L = 4$ . The signal to noise ratio (SNR) is set at 40 dB. Figure 1 shows the temporal representation of the original sources, their convolutive mixtures and the recovered signals by the SOMOD algorithm. For the same experiment, Figure 2 shows the Time Frequency representation of the original sources, their convolutive mixtures and the recovered signals by the SOMOD algorithm. The kernel used for the computation of the TFDs is the Choi-Williams kernel [9]. This example is an illustration of the success of the proposed algorithm in separating two sources.



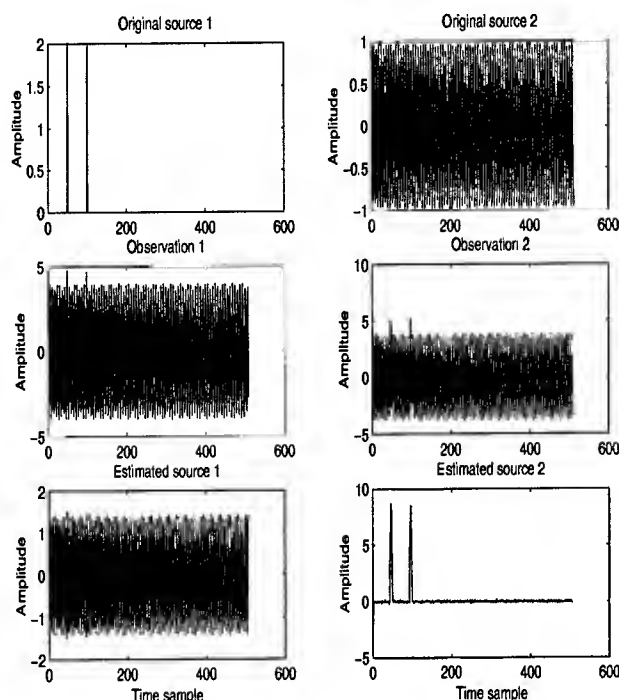


Figure 1: Separation example (Time representation): SNR=40dB.

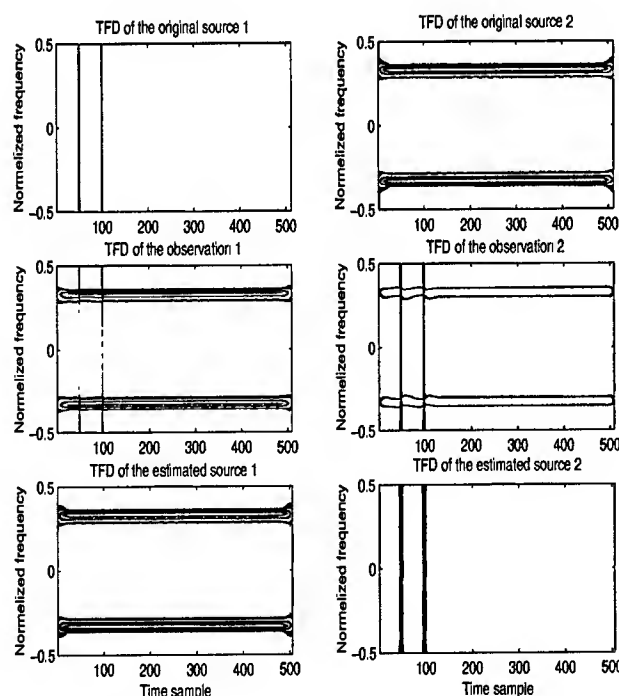


Figure 2: Separation example (Time Frequency representation): SNR=40dB.

## Example 2

We present here a simulation to illustrate the effectiveness of our algorithm in deconvolving speech signals. The parameter settings are :

- $M = 2$ ,  $N = 2$  and  $L = 3$ .
- The two speech signals are sampled at 16kHz.
- Signal to Noise Ratio (SNR) = 10 dB.
- The transfer function matrix of the simulated multi channel is given by,

$$\mathbf{H}(z) = \begin{bmatrix} -0.40 + 0.82z^{-1} + 1.29z^{-2} & 1.19 - 0.02z^{-1} - 1.60z^{-2} \\ 0.69 + 0.71z^{-1} + 0.67z^{-2} & -1.20 - 0.16z^{-1} + 0.26z^{-2} \end{bmatrix}$$

Figure 2 shows the original speech signals, their convolutive mixtures and the recovered speech signals by the SOMOD algorithm.

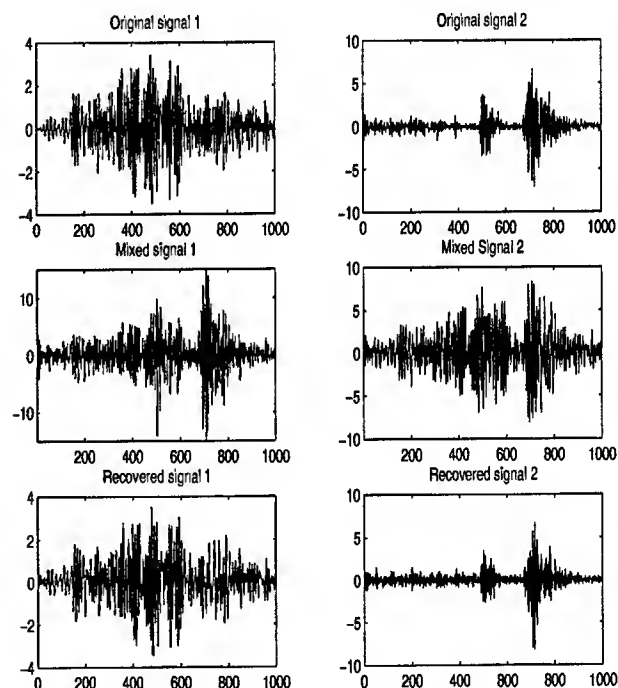


Figure 3: Speech signal separation : SNR=10dB.

## 5. CONCLUSION

In this contribution, we considered the blind deconvolution of MIMO FIR systems driven by mutually decorrelated source signals. We proposed a solution based on the joint diagonalization of spatio-temporal correlation matrices. This technique has been proposed previously

in the case of instantaneous mixtures [1]. An extension of this method to convolutive mixtures has been presented in this paper. It showed to be robust with respect to additive noise. Moreover, it is able to deal with an overestimated source number; since the method provides an  $M \times L$  recovered source subspace instead of the original  $M$  source subspace. A source selection criterion has been defined to select the  $M$  recovered sources among the  $M \times L$  obtained. This method is well suited when applied to the deconvolution of speech signals, which is of great importance in practical applications [7].

## 6. REFERENCES

- [1] A. Belouchrani and K. Abed Meraim and J.-F. Cardoso and E. Moulines, "A blind source separation technique using second order statistics," *IEEE Trans. on SP*, vol. 45, pp. 434-444, Feb. 1997.
- [2] G. B. Giannakis, Y. Inouye, and J. M. Mendel, "Cumulant-based identification of multichannel moving-average processes," *IEEE Trans. on Automat. Contr.*, vol. 34, pp. 783-787, Jul. 1989.
- [3] Y. Inouye and Ruey-wen Liu, "Criteria for direct blind deconvolution of MIMO FIR systems driven by white source signals," in *Proc. ICAS-SP'99*, Phoenix, Arizona, May 1999.
- [4] J. K. Massey and M. K. Sain, "Inverse of linear sequential circuits," *IEEE Trans. Computers*, pp. 330-337, 1968.
- [5] A. Cichocki, S. Amari and J. Cao, "Blind separation of delayed and convolved signals with self-adaptive learning rate," in *In Proc. Nolta'96*, 1996.
- [6] Kari Torkkola, "Blind separation of convolved sources based on information maximization," in *In IEEE Workshop on Neural Networks for Signal Processing, Kyoto, Japan*, Sept. 1996.
- [7] F. Ehlers and H. G. Schuster, "Blind separation of convolutive mixtures and an application in automatic speech recognition in a noisy environment," *IEEE Trans. on SP*, vol. 45, pp. 2608-2612, Oct. 1997.
- [8] G.H. Golub and C.F. Van Loan, *Matrix computations*. The Johns Hopkins University Press, 1989.
- [9] L. Cohen, *Time-frequency analysis*. Prentice Hall, 1995.

# DOA ESTIMATION OF MANY W-DISJOINT ORTHOGONAL SOURCES FROM TWO MIXTURES USING DUET

Scott Rickard<sup>1,2</sup>

Frank Dietrich<sup>2</sup>

<sup>1</sup>Princeton University  
Fine Hall, Washington Road  
Princeton, NJ 08540, USA

<sup>2</sup>Siemens Corporate Research  
755 College Road East  
Princeton, NJ 08540, USA

## 1. ABSTRACT

A novel direction of arrival (DOA) technique is presented which constructs estimates of the relative delay mixing parameters associated with each signal by taking the ratio of time-frequency representations of two mixtures. The technique is based on the Degenerate Unmixing and Estimation Technique (DUET)[1]. If the sources are W-disjoint orthogonal, meaning that only one signal is active in the time-frequency plane at a given time-frequency, then the ratio only depends on the mixing parameters of one source. The ratio can thus be used to generate estimates of the mixing parameters and these estimates can be clustered to determine both the number of sources present in the mixtures and their associated mixing parameters. The method allows for the estimation of the DOA for many sources using only two receive antennas, whereas traditional techniques require  $N$  antennas to estimate  $N - 1$  angles of arrival. Simulation results are presented and compared to MUSIC, ESPRIT, and other DOA estimation techniques.

## 2. INTRODUCTION

The goal of accurately estimating the arrival angle of a signal on an antenna array is long standing in the field of signal processing. Direction of arrival estimation is important for such tasks as tracking the signal emitter and smart antenna array processing for interference reduction in mobile wireless systems.

Most DOA techniques require  $N$  antennas to estimate  $N - 1$  angles of arrival. A notable exception to the  $N - 1$  angles of arrival rule uses forth-order cumulants to estimate three time delays from two mixtures[2]. One advantage of the technique presented here is that it requires only two antenna elements to estimate the arrival angle of an arbitrary number of sources. This reduction in the required number of antenna elements is made by assuming the sources are W-disjoint orthogonal.

This paper applies the work on the Degenerate Unmixing and Estimation Technique(DUET) on W-disjoint orthogonal signals originally proposed in [3] to wireless signals. W-disjoint orthogonal signals have disjoint support for their time-frequency representation. For example, multiple M-ary frequency shift keyed signals are W-disjoint orthogonal, except for the occasional hit when two or more signals transmit at the same frequency at the same time. Another

(perhaps surprising) example of W-disjoint orthogonal signals is speech. Tests show that voice data satisfies the W-disjoint orthogonality constraint closely enough to allow accurate angle of arrival estimation and blind separation[3, 4].

In essence, the W-disjoint orthogonal assumption assumes that all signals are instantaneously separated in the frequency domain. Thus the technique presented herein would not work, for example, when the signals are sinusoids modulated at exactly the same frequency as the signals in that case would not be W-disjoint orthogonal.

Note that one could employ a bank of narrow bandpass filters to create a number of narrowband signal channels and use the DOA estimation schemes described in the above overview literature. In this case, with  $N = 2$ , the standard DOA estimation technique would be able to estimate the angle of arrival of one source per channel. If the source to bandpass channel mapping changes, as would happen rapidly with frequency hopped or voice signals, the multitude of estimates from different channels for different times must be combined in some fashion[5, 6]. One advantage of this technique is that the estimates from different channels are combined inherently as part of the clustering.

Section 3 describes the mixture model, defines W-disjoint orthogonality, and proposes a number of possible angle of arrival estimators based on the model and assumptions. Section 4 presents results of the DOA estimator performance, comparing results with ESPRIT, MUSIC, and other standard DOA techniques.

## 3. MIXING PARAMETER ESTIMATION

### 3.1. Signal Mixing

Consider the measurements of a pair of antenna elements where only the direct path is present. In this case, each mixture  $x_i(t)$  is the sum of delayed attenuated sources signals. We can absorb the attenuation factor and time delay associated with each source to the first antenna element into the definition of the sources and represent mixing in the frequency domain as,

$$\begin{bmatrix} X_1(\omega) \\ X_2(\omega) \end{bmatrix} = \begin{bmatrix} 1 & \dots & 1 \\ a_1 e^{-j\omega\delta_1} & \dots & a_N e^{-j\omega\delta_N} \end{bmatrix} \begin{bmatrix} S_1(\omega) \\ \vdots \\ S_N(\omega) \end{bmatrix}, \quad (1)$$

where  $\delta_i$  is the arrival delay between adjacent array elements resulting from the angle of arrival for source  $i$  and  $\alpha_i$  is the relative attenuation factor for each source between array elements. We denote the maximum possible time delay between array elements as  $\Delta$  and thus  $|\delta_i| \leq \Delta, \forall i$ .

### 3.2. Source Assumptions

Given a windowing function  $W(t)$ , we call two functions  $s_i(t)$  and  $s_j(t)$  **W-disjoint orthogonal** if the supports of the windowed Fourier transforms of  $s_i(t)$  and  $s_j(t)$  are disjoint. The windowed Fourier transform of  $s_i(t)$  is defined,

$$\mathcal{F}^W(s_i(\cdot))(\omega, \tau) = \int_{-\infty}^{\infty} W(t - \tau) s_i(t) e^{-j\omega t} dt, \quad (2)$$

which we will refer to as  $S_i^W(\omega, \tau)$  when appropriate. The W-disjoint orthogonality assumption can be stated concisely,

$$S_i^W(\omega, \tau) S_j^W(\omega, \tau) = 0, \forall i \neq j, \forall \omega, \tau. \quad (3)$$

We will assume, as is common in array processing literature, the physical separation of the sensors is small enough relative to the carrier and bandwidth of the signal such that the relative delay between the sensors can be expressed as a phase shift of the signal[7]. This assumption is known as the **narrowband assumption** in array processing and can be expressed for our purposes as,

$$\mathcal{F}^W(s_i(\cdot - \delta))(\omega, \tau) = e^{-j\omega\delta} \mathcal{F}^W(s_i(\cdot))(\omega, \tau), \forall |\delta| \leq \Delta. \quad (4)$$

### 3.3. Amplitude-Delay Estimation

For W-disjoint orthogonal sources under the narrowband assumption, we note that mixing can be expressed in the time-frequency domain as,

$$\begin{bmatrix} X_1(\omega, \tau) \\ X_2(\omega, \tau) \end{bmatrix} = \begin{bmatrix} 1 \\ \alpha_i e^{-j\omega\delta_i} \end{bmatrix} S_i(\omega, \tau), \text{ for some } i. \quad (5)$$

Due to the sources being W-disjoint orthogonal, mixing for a given  $(\omega, \tau)$  is a function of at most one source. Thus, the mixing parameters can be approximated for a given  $(\omega, \tau)$  using,

$$(\hat{\alpha}_i, \hat{\delta}_i) = \left( \left\| \frac{X_2^W(\omega, \tau)}{X_1^W(\omega, \tau)} \right\|, \Im(\log(\frac{X_2^W(\omega, \tau)}{X_1^W(\omega, \tau)})) / \omega \right), \quad (6)$$

for some  $i$ , where  $\Im$  denotes taking the imaginary part. Equation 6 has been shown to yield accurate mixing parameter estimates for appropriate  $W(t)$  under a variety of noise (independent additive white Gaussian noise) and multipath conditions[3]. Note that for baseband representations of wireless signals, we must divide by  $\omega - \omega_c$  instead of  $\omega$  in Equation 6 where  $\omega_c$  is the carrier frequency.

Using Equation 6, every  $(\omega, t)$  yields an estimate pair for the relative amplitude-delay parameter associated with one source. For W-disjoint orthogonal signals, if we were to calculate amplitude-delay estimates from a number of time-frequency points, we would expect to see clusters around the true delay mixing parameters for each source. Figure 1 shows the estimate clusters for a ten source mixing simulation. If we were to use a standard clustering technique

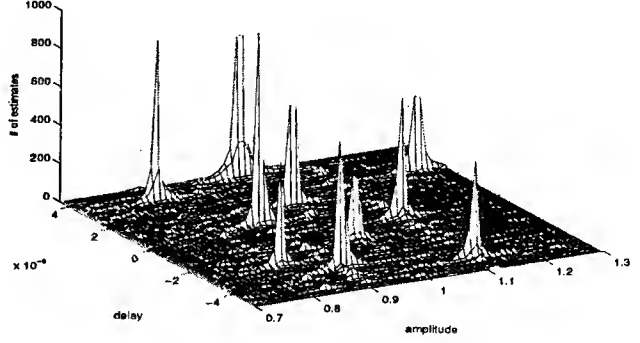


Figure 1: Two-dimensional histogram of number of DUET estimates for delay/amplitude mixing parameters for ten sources obtained using two mixtures. The sources were M-ary FSK wireless signals asynchronously arriving at the two antenna elements. The ten peaks correspond to the correct relative amplitude and delay mixing parameters. The actual relative amplitudes used in the mixing were (1, 1.2, .8, 1, 1.1, .9, 1, 1.2, 1.1, .9) and the corresponding angle of arrivals were (30°, 42°, 54°, 66°, 78°, 90°, 102°, 114°, 126°, 138°). The units of the relative delay is (fractional) samples.

on the amplitude-delay estimates, the number of clusters found would be the estimate of the number of sources, and the cluster centers would be the amplitude-delay estimates associated with each source. The estimated delay can, of course, be translated in to angle of arrival,  $\alpha$ , via,

$$\alpha = \arcsin(\delta_i / \Delta). \quad (7)$$

#### 3.3.1. Equal weight combining

Multiple delay estimates can be combined into an overall delay estimate in order to mitigate the effects of noise and the inaccuracies of the narrowband assumption via,

$$\hat{\delta}_i = \frac{\sum_{(\omega, t) \in \Omega_i} \Im(\log(\frac{X_2^W(\omega, t)}{X_1^W(\omega, t)})) / \omega}{|\Omega_i|}, \quad (8)$$

and multiple relative amplitude estimates associated with the same source can be combined into an overall estimate,

$$\hat{\alpha}_i = \frac{\sum_{(\omega, t) \in \Omega_i} \left\| \frac{X_2^W(\omega, t)}{X_1^W(\omega, t)} \right\|}{|\Omega_i|}, \quad (9)$$

where  $\Omega_i$  is a set of  $(\omega, t)$  points (determined to be associated with the  $i$ th cluster) and  $|\Omega_i|$  is the number of points in the set. The above estimators, Equations 8 and 9, will be referred to as the equal weight estimators because the estimate from each time-frequency point being considered gets equal weight in the overall estimate.

#### 3.3.2. Power weight combining

Rather than weighting each estimate equally in the clustering algorithm, we could weight each estimate by the instantaneous power of the mixture for the time-frequency pair

generating the estimate. Thus, power weighted estimates are,

$$\hat{\delta}_i = \frac{\sum_{(\omega, \tau) \in \Omega_i} \|X_1^W(\omega, \tau)\|^2 \Im(\log(\frac{X_1^W(\omega, \tau)}{X_2^W(\omega, \tau)})) / \omega}{\sum_{(\omega, \tau) \in \Omega_i} \|X_1^W(\omega, \tau)\|^2}, \quad (10)$$

$$\hat{a}_i = \frac{\sum_{(\omega, \tau) \in \Omega_i} \|X_1^W(\omega, \tau)\|^2 \left\| \frac{X_2^W(\omega, \tau)}{X_1^W(\omega, \tau)} \right\|}{\sum_{(\omega, \tau) \in \Omega_i} \|X_1^W(\omega, \tau)\|^2}. \quad (11)$$

### 3.3.3. DOA Product Estimator

An alternative delay estimator can be formed noting,

$$X_1^W(\omega, \tau) \overline{X_2^W(\omega, \tau)} = a_i e^{i\delta_i \omega} \|S_i^W(\omega, \tau)\|^2. \quad (12)$$

Thus, we can estimate the delay parameter via,

$$\hat{\delta} = \Im(\log(X_1^W(\omega, \tau) \overline{X_2^W(\omega, \tau)})) / \omega. \quad (13)$$

It is not possible to estimate the relative amplitude parameter using the product.

## 4. SIMULATION RESULTS

A realistic scenario was defined and simulations performed to test the performance of the DUET algorithm. Simulations were done in MATLAB[8] and detailed comparisons of the DOA results made to ESPRIT and MUSIC[9, 10]. For all simulations, the two subarrays in the uniform linear array for ESPRIT are displaced by one antenna.

For the simulations, a bit stream with 20 kbit/s data rate was transmitted using M-ary frequency shift keying (FSK) with a carrier frequency of 1 GHz. M-ary FSK transmits information via shifting the carrier frequency of a modulated waveform to one of  $M$  values every  $T_s$  seconds. In M-ary FSK, the signal set is defined as,

$$s(t) = \cos(\omega_c + (i-1)\Delta\omega)t, 0 \leq t \leq T_s, i = 1, 2, \dots, M, \quad (14)$$

where  $\Delta\omega = \pi/T_s$ . The  $M$  orthogonal signals are of equal duration and power and are separated by at least  $1/2T_s$  Hz. Multiple M-ary signals generated from independent bit streams are nearly W-disjoint orthogonal, provided the probability that two users transmit in the same frequency bin at the same time is small. The M-ary FSK system had 60 frequency bins with a spacing of 160 kHz. Parameters and signalling method were chosen to model narrowband signalling, as M-ary FSK is a narrowband technique, and also serve as an abstraction for frequency hopped spread spectrum in antenna array systems. The signal sources were asynchronous. Therefore each signal was delayed randomly simulating asynchronously arriving bits. The sources were delayed by choosing random angles of arrival and mixed synthetically assuming the antennas in the uniform linear array were equally spaced with half wavelength separation. For simplicity, all amplitude parameters were set to unity. All simulations were done in the complex baseband representation of the received signals.

Figure 2 shows the histogram of estimates for one experiment with one source at  $-20^\circ$  for the power weighted

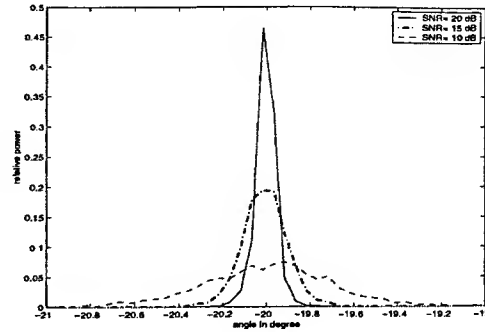


Figure 2: Histogram of angle estimates with the DUET power weighted estimator for one source at  $-20^\circ$  at different noise levels. The estimated variance increases with the noise level.

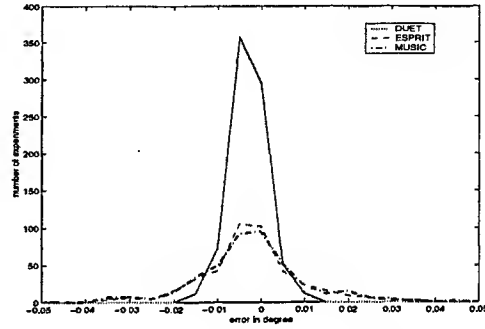


Figure 3: Histogram of estimation errors: one source, no noise, 800 experiments with random angle of arrival, two antennas, comparison of DUET, ESPRIT and MUSIC. DUET has better performance than ESPRIT and MUSIC.

estimator. The power weighted histogram was constructed by summing up the instantaneous power for all the estimates which fall in a given bin. As expected, for lower signal to noise ratios the peak widens and DOA estimates with larger variance are obtained.

We compared the accuracy of the estimates of the DUET algorithm to both ESPRIT and MUSIC. For the DUET algorithm, the ratio form of the estimator was used. The performance of the product form of the estimator was similar to the ratio form. In the simulations with no noise, the equally weighted estimator was used. In the simulations with noise, the power weighted estimator was used. The performance of both estimators was similar.

In the first set of comparison simulations, one source was randomly placed at 800 different angles and the DOA estimated for each case. This is a fair comparison as all three algorithms can perform the estimation with just two antennas. The histogram of absolute errors (Figure 3) for the no noise case shows that DUET outperforms ESPRIT and MUSIC.

In the second set of comparison simulations, 10 sources were used. With 10 sources, DUET requires only two antennas, whereas MUSIC and ESPRIT require at least 11 antennas. The directions of arrival for the sources were randomly chosen between  $30^\circ$  and  $150^\circ$ , which is the typical range of

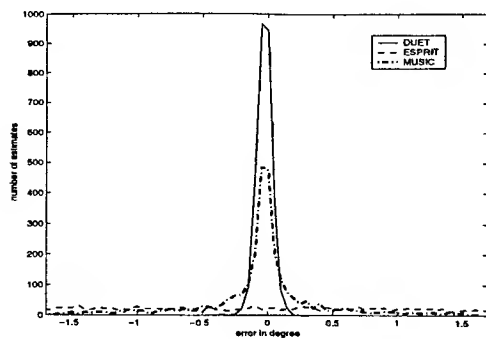


Figure 4: Histogram of estimation errors: ten sources, 20 dB SNR, 300 experiments (random angle of arrival for each source), two antennas for DUET, 15 antennas for MUSIC and ESPRIT. Despite using only two antennas (compared with 15 for the other techniques), DUET estimates the DOA with highest accuracy.

SNR	DUET	ESPRIT	MUSIC	ML	MN
$\infty$ dB	0.013	1.628	1.092	0.050	1.378
20 dB	0.093	1.680	1.088	0.337	1.408
15 dB	0.144	1.681	1.082	0.852	1.408
10 dB	1.484	1.682	1.087	1.445	2.407

Table 1: Maximum absolute error in degree for 90% of estimates. Two antennas were used for DUET, 15 antennas for other algorithms. For example, 90% of DUET's estimates has an error of less than or equal to 0.1444 degree for a SNR of 15 dB. The table shows that DUET has the best performance for  $\infty$  dB, 20 dB, and 15 dB SNR. At 10 dB SNR, DUET has slightly worse performance than MUSIC and ML.

angles when deploying an antenna array in a sectorized cellular communication system. Simulations were performed with 10 sources for no noise, 20 dB, 15 dB, and 10 dB signal to noise ratios. Two antennas were used for DUET, 15 for ESPRIT and MUSIC. The histograms of absolute errors are shown in Figure 4 for the 20 dB case and the tables contain results for all the noise levels and also contain results from the well-known ML and Min-Norm(MN) methods[8]. The results show that DUET has better performance than ESPRIT and MUSIC for  $\infty$  dB, 20 dB, and 15 dB SNR cases. However, the performance of ESPRIT and MUSIC is relatively invariant to noise while DUET's performance decreases with higher noise levels.

## 5. SUMMARY

A method for DOA estimation for an arbitrary number of W-disjoint orthogonal sources from two receive antennas has been presented. Simulations confirm that the technique, DUET, can be used to estimate the DOAs of multiple wireless signals with higher accuracy than MUSIC and ESPRIT for a range of noise levels. The results were obtained for 10 sources using two antennas for the DUET algorithm and 15 antennas for the competing methods.

SNR	DUET	ESPRIT	MUSIC	ML	MN
$\infty$ dB	100 %	50.3 %	79.5 %	98.1 %	72.5 %
20 dB	99.4 %	49.5 %	78.9 %	92.4 %	71.7 %
15 dB	97.3 %	49.7 %	79.0 %	85.3 %	71.7 %
10 dB	75.8 %	49.6 %	79.0 %	68.8 %	71.6 %

Table 2: Percentage of estimates with an absolute error less than 0.5 degree. Two antennas were used for DUET, 15 antennas for other algorithms. For example, 97.3% of DUET's estimates were within .5 degree of the true angle of arrival in the 15 dB case. The table shows that DUET has the best performance for  $\infty$  dB, 20 dB, and 15 dB SNR. At 10 dB SNR, DUET has slightly worse performance than MUSIC.

## Acknowledgements

The authors wish to thank Radu Balan for suggesting the product form of the delay estimator.

## 6. REFERENCES

- [1] A. Jourjine, S. Rickard, and O. Yilmaz. Blind Separation of Disjoint Orthogonal Signals: Demixing N Sources from 2 Mixtures. In *Proc. ICASSP2000, June 5-9, 2000, Istanbul, Turkey*, June 2000.
- [2] B. Emile, P. Comon, and J. Le Roux. Estimation of Time Delays with Fewer Sensors than Sources. *IEEE Trans. on Sig. Proc.*, 46(7):2012-2015, July 1998.
- [3] A. Jourjine, S. Rickard, and Ö. Yilmaz. Blind Separation of Disjoint Orthogonal Sources. Technical Report SCR-98-TR-657, Siemens Corporate Research, 755 College Road East, Princeton, NJ, Sept. 1999.
- [4] <http://www.princeton.edu/~srickard/bss.html>.
- [5] D. Kraus and J. Böhme. Maximum Likelihood Location Estimation of Wideband Sources Using the EM Algorithm. *Proceedings IFAC/ACASP-92, Grenoble, France*, pages 467-471, 1992.
- [6] S. Sivanand and J.-F. Yanf and M. Kaveh. Focusing Filters for Wideband Direction Finding. *IEEE Trans. on Sig. Proc.*, 39:437-445, Feb. 1991.
- [7] H. Krim and M. Viberg. Two Decades of Array Signal Processing Research, The Parametric Approach. *IEEE Signal Processing Magazine*, pages 67-94, July 1996.
- [8] A. Swami, J. Mendel, and C. Nikias. *Higher Order Spectral Analysis Toolbox User's Guide - For Use with MATLAB*. The MathWorks, Inc., 1998.
- [9] R. Roy and T. Kailath. ESPRIT-Estimation of Signal Parameters Via Rotational Invariance Techniques. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 37(7):984-995, July 1989.
- [10] R. Schmidt. Multiple Emitter Location and Signal Parameter Estimation. *IEEE Transactions on Antennas and Propagation*, AP-34(3):276-280, March 1986.

# BLIND SEPARATION OF NON-CIRCULAR SOURCES

Jérôme Galy<sup>1</sup>, Claude Adnet<sup>2</sup>

(1) LIRMM, UMR CNRS 5506  
165 rue Ada, 34392 Montpellier Cedex 05, France

(2) Thomson-Csf.Airsys  
7 rue des mathurins 92223 Bagneux Cedex, France  
galy@lirmm.fr, claude.adnet@airsys.thomson-csf.com

## ABSTRACT

Blind Source Separation is now a well known problem. When a priori informations about the propagation or the geometry of the array are not available, the model can be generalized to a blind source separation model. It supposes the statistical independence of the sources and their non-gaussianity. In this paper, we focus on an algorithm, called Canonical Correlation Analysis, based on the use of second order statistics.

## 1. CANONICAL CORRELATION ANALYSIS

The Canonical Correlation Analysis is a method of treatment which allows to study the correlation between two sets of data.

We can have one set of data  $k$  fonction of the observed signals.

$$k = g[x] \quad (1)$$

In blind source separation, and in particular when we are interested in anti-jamming processing, we divide the received signals into source signals and noise signals. The second set of data  $k$  is get from the observed signals  $x$  of the antenna. This processing is selected to keep the signal of interest :

$$\begin{aligned} x &= x_{interest} + x_{noise} \\ k &= k_{interest} + k_{noise} \\ R_{xk} &= E[xk^H] = E[x_{interest}k_{interest}^H] \end{aligned} \quad (2)$$

The Canonical Correlation Analysis can be divided in several steps.

The first step is to write the two whitened sets of data :

$$\begin{aligned} \Xi_x &= R_x^{-1/2} x \\ \Xi_k &= R_k^{-1/2} k \end{aligned} \quad (3)$$

with :  $E[\Xi_x \Xi_k^H] = R_x^{-1/2} R_{xk} R_k^{-1/2}$   
We can find the eigenvalues of this matrix :

$$E[\Xi_x \Xi_k^H] = U \Sigma^2 V^H \quad (4)$$

If we develop two new matrix :

$$\alpha = U^H \Xi_x \quad (5)$$

and

$$\beta = V^H \Xi_k \quad (6)$$

then we can say that  $\alpha$  has all of the information on  $\Xi_k$  which can be obtained from  $\Xi_x$  and mutually,  $\beta$  has all of the information on  $\Xi_x$  which can be obtained from  $\Xi_k$ .

So we resolve  $E[\alpha\beta^H] = \Sigma^2$  under :  $E[\alpha\alpha^H] = I = E[\beta\beta^H]$ .

Suppose that we have two sets of data  $x$  and  $k$  available.

The Canonical Correlation Analysis consists in defining two matrix  $W_x$  and  $W_k$  in order to  $W_x^H x$  and  $W_k^H k$  must be the more correlated.

So we have :

$$\begin{aligned} \alpha &= W_x^H x \\ \beta &= W_k^H k \end{aligned} \quad (7)$$

The Canonical Correlation Analysis minimizes the criterion :

$$\Phi(W_k, W_x) = E[|W_k^H k - W_x^H x|^2] \quad (8)$$

under :

$$W_k^H R_k W_k = 1 \quad (9)$$

and

$$W_x^H R_x W_x = 1 \quad (10)$$

The minimization can be written :

$$\Phi(W_k, W_x) = \text{trace} \left\{ \begin{aligned} &W_k^H R_k W_k + W_x^H R_x W_x - \\ &W_k^H R_k x W_x - W_x^H R_x k W_k \end{aligned} \right\} \quad (11)$$

To minimize  $\Phi(W_k, W_x)$ , we must derive from each component of  $W_k, W_x$  and use Lagrange operations  $\Lambda$  et  $\Delta$  which are not discussed in our paper..

We have now two equations :

$$R_{xk} W_k = R_x W_x \Lambda \quad (12)$$

$$R_{kx} W_x = R_k W_k \Delta \quad (13)$$

We modify the equation, (multiplication by  $R_{kx}^{-1} R_{kx}$ ) we can see :

$$\begin{aligned} R_{xk} W_k &= R_x R_{kx}^{-1} R_{kx} W_x \Lambda \\ &= R_x R_{kx}^{-1} R_k W_k \Delta \Lambda \end{aligned} \quad (14)$$

If we are interested in  $W_x$ , we can have the dual equation.

If we call  $T = \Delta \Lambda$ , then :

$$R_k^{-1} R_{kx} R_x^{-1} R_{xk} W_k = W_k T \quad (15)$$

If we multiply by  $R_k^{1/2}$ , we have the equation :

$$R_k^{-1/2} R_{kx} R_x^{-1/2} R_x^{-1/2} R_{xk} R_k^{-1/2} R_k^{1/2} W_k = R_k^{1/2} W_k T \quad (16)$$

With  $R_k^{-1/2} R_{kx} R_x^{-1/2} = D$ , then :

$$D D^H \tilde{W}_k = \tilde{W}_k T \quad (17)$$

with  $\tilde{W}_k = R_k^{1/2} W_k$

So we can find the eigenvalues and eigenvectors of  $D$ . We choose the  $L$  eigenvectors  $U_1$  corresponding to  $L$  higher eigenvalues.

The matrix  $W_k$  is now :

$$W_k = R_k^{-1/2} U_1 \quad (18)$$

The argument is the same if we are interested in the matrix  $W_x$ .

## 2. APPLICATIONS

We take the model without noise :  $x = As$ . The matrix  $A$  have the SVD :  $A = U \Sigma V$

### 2.1. SOBI

If the second sets of data can be deduced from the first ones with the addition of a delay on the signal :

$$k = As(t - \tau) \quad (19)$$

and

$$x = As(t) \quad (20)$$

We can write :  $R_x = R_k = U \Sigma^2 U^H$ .

We can have :

$$\Xi_x = R_x^{-1/2} x = Vs(t) \quad (21)$$

$$\Xi_k = R_k^{-1/2} k = Vs(t - \tau)$$

with :  $R_x^{-1/2} = R_k^{-1/2} = \Sigma^{-1} U^H$

We have also :

$$E[\Xi_x \Xi_k^H] = V R_s(\tau) V^H \quad (22)$$

To specify  $V$ , *Belouchrani* in SOBI [1] choose to make a joint diagonalization of a set of matrix using second order statistics. This approach can be compared with the *Cardoso* and *Souloumiac* method for the *Jade* algorithm [2].

The estimated  $V$  allows to form the estimated mixing matrix  $\hat{A}$  :

$$\hat{A} = R_x^{1/2} V \quad (23)$$

and the estimated outputs are :

$$\hat{s}(t) = V^H R_x^{-1/2} x \quad (24)$$

### 2.2. Non-circular Source Separation

In our case, the signals (*BPSK*) are non-circular and the noise signals are circular. If the interference signals are  $j(t)$  and the *BPSK* are  $s(t)$ , we can see that :

$$E[s(t)^2] \neq 0 \quad (25)$$

and

$$E[j(t)^2] = 0 \quad (26)$$

The signals (*BPSK*) are non-circular, if we want to eliminate the circular interferences, we can use for  $k(t)$  the conjugate of  $x(t)$  :

$$k(t) = x(t)^* \quad (27)$$

The model is always  $x = As$  :

$$k(t) = A^* s(t)^* \quad (28)$$

We look at the matrix  $A$  which can be divided in eigenvalues and we can write the conjugate  $A$  noted  $A^*$  :

$$A^* = U^* \Sigma V^* \quad (29)$$



The correlation matrix of the source signals is :

$$R_x = E [x(t)x(t)^H] = U\Sigma^2U^H \quad (30)$$

Now the correlation matrix of the set of data  $k(t)$  can be written :

$$R_k = E [k(t)k(t)^H] = E [x(t)^*k(t)^T] = U^*\Sigma^2U^T \quad (31)$$

If we have the whitened sets of data :

$$\Xi_x = R_x^{-1/2}x = \Sigma^{-1}U^Hx = Vs(t) \quad (32)$$

$$\Xi_k = R_k^{-1/2}k = \Sigma^{-1}U^Tx = V^*s^*(t)$$

with :

$$R_x^{-1/2} = \Sigma^{-1}U^H \quad (33)$$

and

$$R_k^{-1/2} = \Sigma^{-1}U^T \quad (34)$$

We have :

$$E [\Xi_x \Xi_k^H] = V E[ss^T]V^T \quad (35)$$

The matrix  $E[ss^T]$  only contains the informations on non-circular signals. The SVD factoring of  $E[ss^T]$  allows to estimate  $V$  and to find only the non-circular signals of the mixing.

Th estimation of  $V$  allows to have the estimated mixing matrix  $\hat{A}$  :

$$\hat{A} = R_x^{1/2}V \quad (36)$$

This research of the mixing matrix can be qualified a *blind separation* because no information on antenna, on propagation or on signals is necessary to have the 'filter'. The noise signals must be circular to be rejected by this algorithm [3].

One of the applications of this algorithm is the subject of a patent registered with Thomson-CSF.

### 3. RESULTS

#### 3.1. Adaptive Antenna

The antenna is an *MSLC (Multiple Sidelobe Canceller) antenna*, which means that we can have one main antenna and some auxiliary elements. Indeed, for the supervision of some particular space areas, we use this kind of antenna which allows to focus on the main antenna the information on the source signal while supervising areas likely to have some jamming signals.

If we consider the sectional elevation, we can recognize on figure 1 the main antenna and an auxiliary element. The main antenna has a constant value (3-4

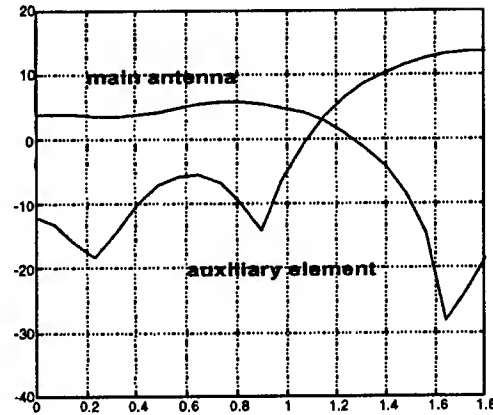


Figure 1: MSLC Antenna

dB) between  $0^\circ$  and  $1^\circ$  (angle of sight) and the auxiliary element has some variations between these angles of sight.

The source signal and the jamming sources are :

- 1 source signal located at  $0^\circ$  (angle of sight),  $0^\circ$  (yaw angle) and with power 20dB.
- 2 gaussian jammers located one at  $0^\circ$  (angle of sight) and  $1.5^\circ$  (yaw angle) with power 20dB and other at  $1.7^\circ$  (angle of sight) and  $0^\circ$  (yaw angle) with power 20dB.
- gaussian noise with power 0dB.

This kind of situation is good for the classical treatment (anti-jamming with MSLC). The source signal is located in the sidelobe of the main antenna and the two jammers are located in the middle of the auxiliary elements. Now we can compare the performances of the different algorithms.

#### 3.2. Performances

It's necessary to study the performances of the two algorithms (MSLC and Canonical Correlation Analysis) when one of the jammers will move and its power will change.

When one of the jammers moves, the performances of the different algorithms can be evaluated :

- keeping the two others fixed.
- changing the last jammer initially located at  $0^\circ$  (yaw angle) along the decreasing angles of sight (variation from  $1.7^\circ$  to  $0^\circ$ ).

The power of the moving jammer is fixed to 20dB and increases to 50dB.

#### – MSLC traitement

On figure 2, lots of elements allow us to verify :

- the Signal to Noise and Interferences Ratio (SINR) becomes weak when the jammer comes near the source signal.
- the SINR becomes weak when the power of the jammer decreases.

These two observations are easily explainable.

For the first one, this is inevitable whichever algorithms we use. This waste of SINR is predictable : if we look at the figure 1, we can see the jammer entering in the main antenna from 1° (angle of sight), and the SINR variation follows the auxiliary element.

The second observation is the result of the **self-jamming of the source signal**. In fact, when the power of the jammer signal is weak, the MSLC algorithm takes the source signal as a jammer and it tries to eliminate it. That is why we call this, self-jamming.

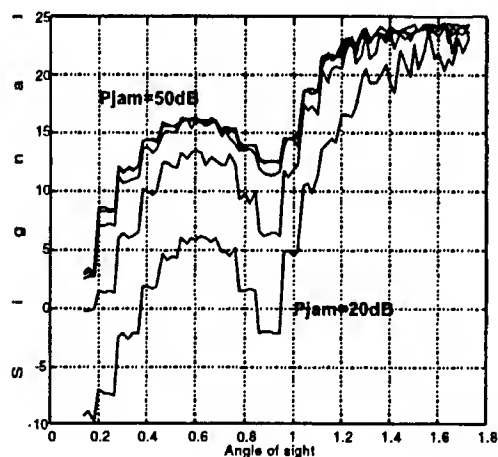


Figure 2: SINR with MSLC Algorithm

#### – Canonical Correlation Analysis

If we make the same experience with the Canonical Correlation Analysis, we can see on figure 3 that the Signal to Noise and Interferences Ratio does not depend on the power of the jammer. The self-jamming of the source signal has disappeared. Whichever the jammer power, the SINR only depends on the auxiliary element.

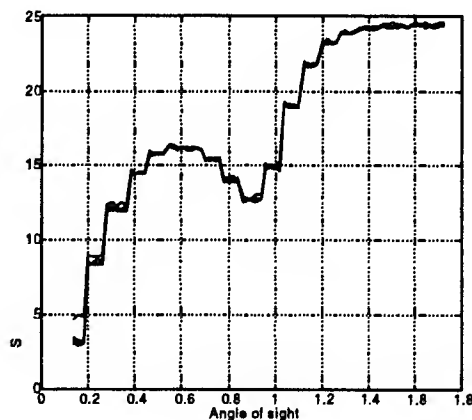


Figure 3: SINR with Canonical Correlation Analysis

## 4. CONCLUSION

If we can have specific information on the source signals, it is better to use methods only based on the use of second-order statistics. These methods are less expensive for the calculation than higher order statistics.

The BPSK signals are non-circular while the jammer sources are gaussian and circular. For Separation Sources, the best technique is the Canonical Correlation Analysis, the results show that this method is effective to avoid the self-jamming of the source signal.

In the case of a MSLC antenna, the performances with Correlation Canonical Analysis are the same as these with JADE algorithm [2] using higher-order statistics [3].

## 5. REFERENCES

- [1] A. Belouchrani 'Séparation autodidacte de sources', Phd Thesis ENST Paris, 1995.
- [2] J.F. Cardoso and A. Souloumiac 'An Efficient Technique for Blind Separation of Complex Source', Proceedings IEEE SP Workshop on Higher-Order Statistics, Lake Tahoe USA Juillet 1993, pp: 275-279.
- [3] J. Galy 'Antenne adaptative : du second ordre aux ordres supérieurs. Applications aux signaux de télécommunications', Phd Thesis University Paul Sabatier Toulouse, Avril 1998.

# BLIND IDENTIFICATION OF SLIGHTLY DELAYED MIXTURES

Gilles Chabriel and Jean Barrère

MS/GESSY - ISITV  
Université de Toulon et du Var  
Av. Georges Pompidou, BP 56  
83162 LA VALETTE DU VAR CEDEX (FRANCE)  
Fax: 33 494 142 598  
chabriel@isitiv.univ-tln.fr  
barrere@isitiv.univ-tln.fr

## ABSTRACT

In this work, we are interested in the separation of  $N$  propagating source signals recorded simultaneously by a set of receivers. To solve this "cocktail-party problem", we propose to collect a set of spatially close microphones. On each sensor, signals are received with the same attenuation but with different time delays. The linear memoryless conventional model for source separation is then no more suitable. However, when time delays are small in comparison with the coherence time of each source, we show that this problem can be simplified building up a particular set of instantaneous mixtures involving derivatives of sources with respect to time. Then, sources can be extracted using second-order methods. The limitations of the method are exposed : we explain what a small delay is and we show that the number of sources can't exceed 3. The validity of the proposed approach is confirmed by computer simulations. Finally, we apply our method to an experiment where two source signals are extracted from their mixtures observed with two omnidirectional microphones in a normal room.

## 1. BASIC ASSUMPTIONS AND MODEL

Let consider  $N$  sources assumed to be statistically independent, localized and differently colored, propagating in an echo-free environment. These sources are recorded by a set of  $M$  sensors spatially close one to the others. We also assume that sources are far from sensors, so the propagation model can be approximate as a far field model and the power of the contribution of one source on each sensor is the same.

In this work, the presence of additive noise on observations will not be treated. For simplicity, we will expose only the case where there is the same number

of observations than the number of sources.

Because of the proximity of sensors, we assume that the contribution of each source received by sensors are the same except a relative propagation delay from one sensor to the other.

Denoting  $\mathbf{y}_s = [y_{s1}(t), y_{s2}(t), \dots, y_{sN}(t)]^T$  the observation vector and  $\mathbf{x} = [x_1(t), x_2(t), \dots, x_N(t)]^T$ , the sources contribution vector, we can write:

$$\begin{aligned} y_{s1}(t) &= x_1(t) + x_2(t) + \dots + x_N(t) \\ y_{si}(t) &= \sum_{j=1}^{j=N} x_j(t - \tau_{i,j}), \quad i = 2, \dots, N, \end{aligned} \quad (1)$$

where  $\tau_{i,j}$  represents the relative delay of source  $x_j(t)$  observed on the  $i^{th}$  sensor versus the first observation  $y_{s1}(t)$ .

Let's consider a contribution  $x_i(t - \tau_{i,j})$  from system (1). Its Fourier Transform (FT) is:

$$\text{FT}[x_j(t - \tau_{i,j})] = X_j(\nu) e^{-2i\pi\nu\tau_{i,j}},$$

where  $\nu$  represents the frequency variable.

The Taylor expansion of  $e^{-2i\pi\nu\tau_{i,j}}$  is  $\sum_{k=0}^{k=\infty} \frac{(2i\pi\nu\tau_{i,j})^k}{k!}$ .

In a physical context, the sensors have a given bandwidth  $[\nu_{min}, \nu_{max}]$ . Let's denote  $\nu_M$  ( $\nu_M < \nu_{max}$ ) the maximum frequency such as  $X_j(\nu) \neq 0$  for all  $j$ .

In case of compact sensor array, we assume that the delays are slight such as:

$$\tau_{i,j}^2 \ll \frac{1}{2\pi^2\nu_M^2}, \quad \forall i, j.$$

Then, in the previous Taylor expansion we neglect terms higher than order one.

So, we can give the following approximation for the observations  $y_{si}(t)$ ,  $i = 2 \dots N$ :

$$\begin{aligned} y_{si}(t) \approx & x_1(t) - \tau_{i,1}\dot{x}_1(t) + \\ & x_2(t) - \tau_{i,2}\dot{x}_2(t) + \\ & \dots + \\ & x_N(t) - \tau_{i,N}\dot{x}_N(t). \end{aligned} \quad (2)$$

where  $\dot{x}_i(t)$  expresses the first derivative of  $x_i(t)$ . Introducing  $y_1(t) = \dot{y}_{s1}(t)$  the first derivative of first observation, and denoting  $y_i(t) = y_1(t) - y_{si}(t)$ , with  $i = 2 \dots N$ , we obtain the following system:

$$\begin{aligned} y_1(t) &= \dot{x}_1(t) + \dots + \dot{x}_N(t) \\ y_i(t) &\approx \tau_{i,1}\dot{x}_1(t) + \dots + \tau_{i,N}\dot{x}_N(t), \quad i = 2 \dots N \end{aligned} \quad (3)$$

which can be rewritten in vector and matrix notations as:

$$\mathbf{y}(t) = \mathbf{M}\dot{\mathbf{x}}(t), \quad (4)$$

$$\text{with } \mathbf{M} = \begin{bmatrix} 1 & \dots & 1 & \dots & 1 \\ \tau_{2,1} & \dots & \tau_{2,j} & \dots & \tau_{2,N} \\ \vdots & & \vdots & & \vdots \\ \tau_{N,1} & \dots & \tau_{N,j} & \dots & \tau_{N,N} \end{bmatrix}.$$

The slightly delayed mixture appears now as an instantaneous mixture of derivatives sources.

In (4),  $\mathbf{M}$  is the unknown memoryless mixing matrix.  $\mathbf{M}$  is a  $N \times N$  matrix assumed to be full column rank (this assumption will be discussed later).

**Remark:**

If we have one sensor more than sources, the derivation of the reference observation can be avoided. In this case the mixing matrix becomes :

$$\mathbf{M} = \begin{bmatrix} \tau_{2,1} & \dots & \tau_{2,j} & \dots & \tau_{2,N} \\ \vdots & & \vdots & & \vdots \\ \tau_{N+1,1} & \dots & \tau_{N+1,j} & \dots & \tau_{N+1,N+1} \end{bmatrix}.$$

## 2. IDENTIFICATION OF INSTANTANEOUS LINEAR MIXTURES

Because of the spectra differences of sources, the problem can be solved by any classical blind identification

method for instantaneous mixtures using second-order statistics of the observations (see Tong's AMUSE [1] [2], SOBI [3], IMISO [5] or [4] ...).

The blind identification problem consists in estimating a separating matrix  $\mathbf{S}$  such as:  $\mathbf{S}\mathbf{M} = \mathbf{D}\mathbf{P}$ , where  $\mathbf{D}$  is a regular diagonal matrix,  $\mathbf{P}$  is a permutation matrix.

The product of  $\mathbf{S}$  with the observations leads to:

$$\mathbf{z}(t) = \mathbf{D}\mathbf{P}\dot{\mathbf{x}}(t),$$

representing the sources derivatives except for one permutation and a scaling factor.

Most of second order methods are based on the diagonalization of two differently linearly filtered covariance matrices of the observations.

Consider the spatial covariance matrix of the observations for any delay  $\tau$ :  $\mathbf{R}_{yy}(\tau) = \mathbf{E}[\mathbf{y}(t)\mathbf{y}^T(t+\tau)]$ . From (1), we can write a relation between  $\mathbf{R}_{yy}(\tau)$  and  $\mathbf{R}_{\dot{x}\dot{x}}(\tau)$ , the spatial covariance matrix of the derivative of sources :

$$\mathbf{R}_{yy}(\tau) = \mathbf{M}\mathbf{R}_{\dot{x}\dot{x}}(\tau)\mathbf{M}^T. \quad (5)$$

Mutual independence of sources implies  $\mathbf{R}_{\dot{x}\dot{x}}(\tau)$  to be a diagonal matrix.

Let's linearly filter with the impulse response  $h(t)$  each member of expression (5):

$$(h * \mathbf{R}_{yy})(\tau) = (h * [\mathbf{M}\mathbf{R}_{\dot{x}\dot{x}}\mathbf{M}^T])(\tau).$$

Because the convolution product is linear, it comes:

$$\mathbf{R}^h(\tau) = \mathbf{M}[\mathbf{D}^h(\tau)]\mathbf{M}^T, \quad (6)$$

where  $\mathbf{R}^h(\tau) = (h * \mathbf{R}_{yy})(\tau)$  and  $\mathbf{D}^h(\tau) = (h * \mathbf{R}_{\dot{x}\dot{x}})(\tau)$ .

Because matrix  $\mathbf{R}_{yy}(0)$  is regular, we can introduce the following matrix:

$$\mathbf{R} = [\mathbf{R}_{yy}(0)]^{-1} \mathbf{R}^h(0).$$

Then from (5) and (6), it comes:

$$\mathbf{R} = [\mathbf{M}^T]^{-1} [\mathbf{R}_{\dot{x}\dot{x}}(0)]^{-1} [\mathbf{D}^h(0)] \mathbf{M}^T.$$

We can show that  $(\mathbf{M}^T)^{-1}$  can be estimated except for one diagonal matrix and one permutation matrix from the eigenvector matrix of  $\mathbf{R}$ .

### 3. LIMITATION OF THE METHOD

In previous section we assumed the matrix  $\mathbf{M}$  to be full column rank. Under the assumptions made on the field of sources, the relative delays  $\tau_{ij}$  only depend on the distance between reference sensor and  $i^{th}$  sensor as illustrated in Figure 1:

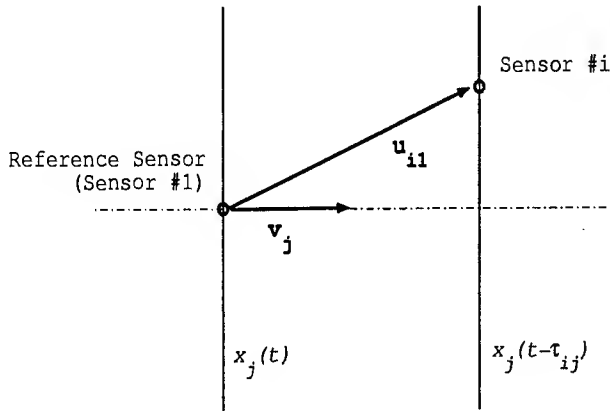


Figure 1: Source #j plane wave

where  $\mathbf{v}_j$  is the unit vector of the plane wave direction of the  $j^{th}$  source and  $\mathbf{u}_{i1}$  is the position vector of the  $i^{th}$  sensor versus sensor reference. The relative delay  $\tau_{ij}$  is given by the scalar product:

$$\tau_{ij} = \frac{1}{c} \mathbf{u}_{i1}^T \mathbf{v}_j, \quad (7)$$

where  $c$  is the propagation velocity supposed to be constant.

It follows that in the three dimensional physical space:

$$\mathbf{M} = \frac{1}{c} [\mathbf{u}_{1i}^T \mathbf{v}_j], \quad i = 2, \dots, N; \quad j = 1, \dots, N;$$

is a non regular matrix for  $N > 3$ . In other words, the identification of such mixture is not possible for more than three sources.

### 4. APPLICATIONS

We present results obtained choosing  $h(t)$  such as a second-order differentiator filter [5].

#### 4.1. Numerical Simulations

Three unit-power synthetic signals are mixed with small delays ( $\tau_{max} \ll (\sqrt{2}\pi\nu_{max})^{-1}$  where  $\nu_{max}$  is the maximum frequency of the observations).

For the particular case presented here (see the top of Figure 2), the synthetic data are obtained by bandpass FIR filtering of white signals. From the spatial location of sensors and directions of arrival of sources we deduct the relative delays using equation (7). The simulated waves are propagating in the air with a velocity equal to  $340 \text{ m sec}^{-1}$ .

The observations are constructed from delayed sources using spectral interpolation. The maximum frequency generated is  $\nu_{max} = 5\text{kHz}$  for a  $11 \text{ kHz}$  sample rate, and the maximum of delays generated is  $\tau_{max} = 21\mu\text{s}$  with a  $3 \text{ cm}$  distance inter-sensors.

On the bottom of Figure 2 we plot the PSD of two of the three observations in order to illustrate that the Power Spectral Densities of observations are identical.

The power spectral densities of estimated sources are plotted in Figure 3.

The performance of our method is measured using the criterion introduced by Shobben and al in [7]. The quality of separation of the  $j^{th}$  separated output is defined as:

$$S_j = 10 \log \left( \frac{\mathbb{E}[(z_{j,x_j})^2]}{\mathbb{E}[(\sum_{i \neq j} z_{j,x_i})^2]} \right),$$

where  $z_{j,x_i}$  is the  $j^{th}$  output when only  $x_i$  is active.

For our numerical experiments the performance measures can be found in the following table:

Estimated Source #1	$S_1 = 17\text{dB}$
Estimated Source #2	$S_2 = 9\text{dB}$
Estimated Source #3	$S_3 = 27\text{dB}$

Better results could be obtained using spatially closer sensors but it conducts to non feasible configurations.

#### 4.2. Real Data

We test the method on real signals recorded by J.T.Ngo et al. [6]. The signals are obtained by two omnidirectional microphones mounted  $1\text{cm}$  apart, recording two human speakers, each  $1\text{m}$  away from sensors. A sample rate of  $22050 \text{ Hz}$  was used for each signal. A comparison with Ngo et al results is plotted of the bottom side of Figure 4.

### 5. EXTENSIONS - CONCLUSIONS

We showed that second-order Blind Separation algorithms can be used to extract propagating colored sources recorded on a compact set of sensors. We have seen that when the source contributions are recorded with the same attenuation on each sensor, the method is

geometrically limited to the extraction of only three sources. When the attenuations are different from one source to each sensor, this limitation vanishes.

The case of noise corrupted sensors requires more sensors than observations and can be treated by classical method as exposed in [2].

The method can be extended for higher delays implying second order development in Taylor series. Some modifications of the second order identification method are necessary.

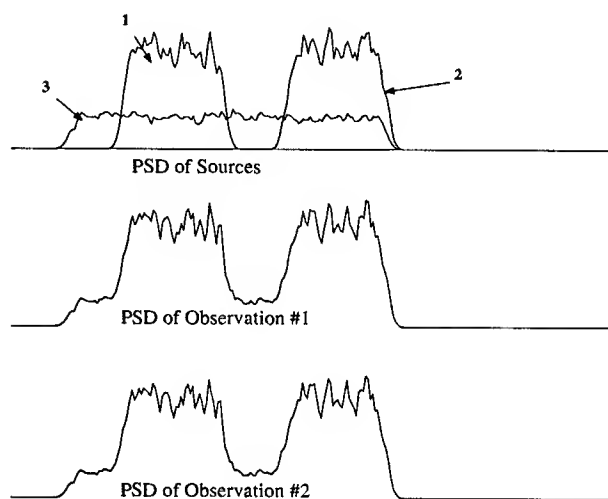


Figure 2: Synthetic Data: sources and observations

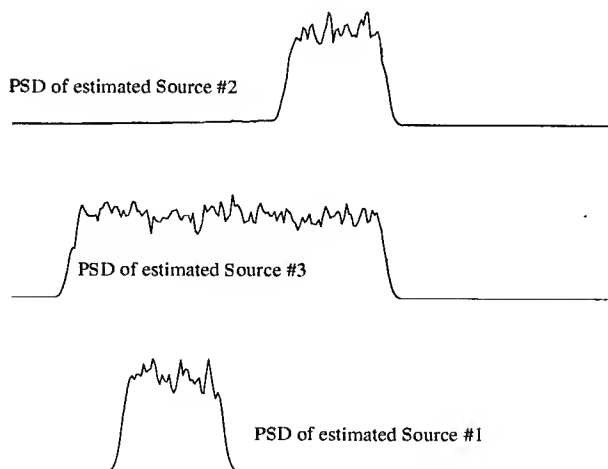


Figure 3: Synthetic Data: estimations

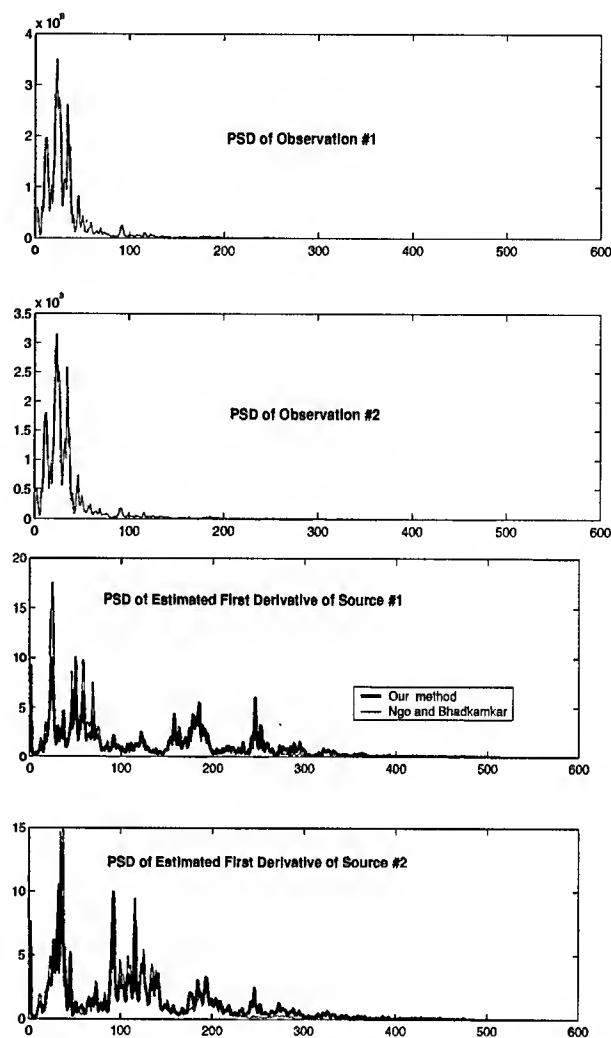


Figure 4: Real Data

## REFERENCES

- [1] L. Tong, V.C. Soon, Y.F. Huang, and R. Liu. *AMUSE: A new blind identification algorithm*, in Proc. 1990 IEEE ISCAS, New Orleans, LA., May 1990.
- [2] L. Tong, R. Liu, and V.C. Soon. *Indeterminacy and identifiability of blind identification*, in IEEE Transactions on Circuits and Systems, Vol. 38, No 5, May 1991.
- [3] A. Belouchrani, K. Abed-Meraim, and J.-F. Cardoso. *A blind source separation technique using second order statistics*, in IEEE Transactions on Signal Processing, Vol. 45, No 2, February 1997.

- [4] L. Fety, *Méthodes de traitement d'antenne adaptées aux radiocommunications*, Doctorat ENST, 1988.
- [5] J.F. Cavassilas, B. Xerri and G. Chabriel. *Séparation autodidacte de sources temporellement corrélées (mélange instantané)*, in GRETSI Symposium, Vol. 1, pp 107-110, Sept. 1997.
- [6] J.T. Ngo, and N.A. Bhadkamkar. *Adaptative Blind Separation of Audio Sources by a Physically Compact Device Using Second Order Statistics*, in ICA'99 First International Workshop on Independant Component Analysis and Signal Separation, pp 256-260, January 11-15, 1999.
- [7] D. Schobben, K. Torkkola, and P. Smaragdis. *Evaluation of blind signal separation methods*, in ICA'99 First International Workshop on Independant Component Analysis and Signal Separation, pp 261-266, January 11-15, 1999.

# ROBUST SOURCE SEPARATION USING RANKS

Lan Xiang, Yinglu Zhang and Saleem A. Kassam

Department of Electrical Engineering  
University of Pennsylvania  
Philadelphia, PA 19104  
E-mail: kassam@ee.upenn.edu

## ABSTRACT

Robustness against deviations from nominal source pdf assumptions is very desirable in blind source separation (BSS) algorithms. In this paper, a new approach for robust BSS is proposed. We modify the EASI (equivariant adaptive separation by independence) algorithms to use ranks of observed signals. Two different methods for evaluation of ranks have been introduced in this paper. Our modified algorithm can be applied to both real-valued data and complex-valued data. Design guidelines are discussed for the nonlinear rank weighting functions in the modified algorithm. Simulation results and some examples are given, showing very good performance.

## 1. INTRODUCTION

Blind Source Separation is the process of recovering a set of independent signals when only mixtures with unknown coefficients are observed. It is usually assumed that little is known about the original sources except that they are mutually independent. Many important theories and applications have been investigated in BSS and more generally in Independent Component Analysis (ICA) [1], [2], [3]. However, little has been done on the robustness issue in BSS. Robustness against deviations from nominal source probability density function (pdf) assumptions is very desirable in BSS algorithms. Some aspects of performance approximation, and the robustness of the EASI (equivariant adaptive separation by independence) algorithms were considered in [4] and [5]. It was shown by Cardoso and Laheld in [2] that the optimum nonlinear function in the EASI algorithms depends on the pdf's of the original sources. Therefore, the performance of the original algorithms is affected by the accuracy of our knowledge on source densities. The robustness of these BSS algorithms was achieved in [4] by using saturating nonlinear functions in the original algorithms. The nature of optimum quantizer

nonlinearities was also studied in [4]. The approach using ranks to improve the robustness of BSS algorithms was first proposed in [5], where the EASI algorithms were modified to use ranks of observed signals. Simulation results in [4], [5] and this paper show that the EASI algorithms fail in estimating the mixing channel when there are deviations from nominal source pdf assumptions.

In this paper, two ranking methods are introduced. Our method using ranks to achieve the robustness of the EASI algorithms can be applied to real-valued data or complex-valued data. Simulation results show good performance with ranks in BSS.

## 2. BLIND SOURCE SEPARATION

The block diagram in Fig. 1 shows a general adaptive BSS scheme for the standard model of instantaneous additive sources.

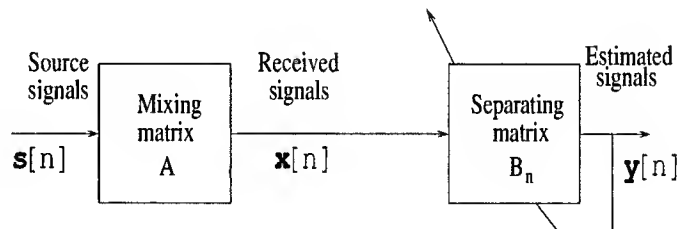


Figure 1: Adaptive BSS of Instantaneous Additive Mixtures

The received discrete-time signal model is that of an  $m$ -dimensional time series  $\mathbf{x}[n] = (x_1[n] \cdots x_i[n] \cdots x_m[n])^T$  of the form:

$$\mathbf{x}[n] = \mathbf{A}\mathbf{s}[n] \quad (1)$$

where

$$\mathbf{s}[n] = (s_1[n] \cdots s_i[n] \cdots s_k[n])^T$$



The channel characteristic between  $\mathbf{s}[n]$  and  $\mathbf{x}[n]$  is defined by the constant mixing matrix  $\mathbf{A}$  of size  $m \times k$ ; there are  $k$  sources and  $m$  receivers. Here we require  $m \geq k$ , which means the number of receivers should be no less than the number of sources.

The objective is to get a separating matrix such that  $\mathbf{y}[n]$  consists of individually scaled and possibly permuted versions of  $\mathbf{s}[n]$ :

$$\begin{aligned}\mathbf{y}[n] &= \mathbf{B}_n \mathbf{x}[n] = \mathbf{B}_n \mathbf{A} \mathbf{s}[n] \\ &= \mathbf{C}_n \mathbf{s}[n]\end{aligned}\quad (2)$$

where

$$\mathbf{C}_n \stackrel{\text{def}}{=} \mathbf{B}_n \mathbf{A} \quad (3)$$

The matrix  $\mathbf{B}_n$  is called the separating matrix after the  $n$ -th iteration. Ideally, for  $n$  large enough,  $\mathbf{B}_n$  has converged to a matrix  $\mathbf{B}$ , and  $\mathbf{C} = \mathbf{B}\mathbf{A}$  is very close to an identity matrix; more generally,  $\mathbf{C}$  is a permutation matrix with arbitrary scaling for each output.

The normalized Equivariant Adaptive Source Separation (EASI) algorithm[2] that we base our modifications on has the form:

$$\begin{aligned}\mathbf{B}_{n+1} &= \mathbf{B}_n - \lambda_n \left[ \frac{\mathbf{y}[n]\mathbf{y}[n]^T - \mathbf{I}}{1 + \lambda_n \mathbf{y}[n]^T \mathbf{y}[n]} \right. \\ &\quad \left. + \frac{\mathbf{g}(\mathbf{y}[n])\mathbf{y}[n]^T - \mathbf{y}[n]\mathbf{g}^T(\mathbf{y}[n])}{1 + \lambda_n |\mathbf{y}[n]^T \mathbf{g}(\mathbf{y}[n])|} \right] \mathbf{B}_n\end{aligned}\quad (4)$$

where  $\lambda_n$  is the adaptation step size, and  $\mathbf{g}(\cdot)$  is a component-wise nonlinear odd function for which design guidelines are available. The term  $(\mathbf{y}[n]\mathbf{y}[n]^T - \mathbf{I})$  in (4) has the effect of driving the diagonal elements of  $\mathbf{C}_n$  to all ones. Meanwhile, the other term  $(\mathbf{g}(\mathbf{y}[n])\mathbf{y}[n]^T - \mathbf{y}[n]\mathbf{g}^T(\mathbf{y}[n]))$  in (4) drives the off-diagonal elements of  $\mathbf{C}_n$  to zeros. Another version of the EASI algorithm is called the original EASI algorithm, which does not have the normalization factors  $(1 + \lambda_0 \mathbf{y}[n]^T \mathbf{y}[n])$  and  $(1 + \lambda_0 |\mathbf{y}[n]^T \mathbf{g}(\mathbf{y}[n])|)$ . The two normalization factors in the normalized EASI algorithm in (4) were introduced to improve the stability of the algorithm. Our simulations in section 5 show the performance of the original/normalized EASI algorithms.

Assuming all the original sources are identically distributed with a differentiable probability density function  $f(s)$ , the optimum  $g$  function in (4) is given in [2] as

$$g_{\text{opt}}(s) = \frac{g_{\text{LO}}(s)}{E g_{\text{LO}}^2(s) - 1} \quad (5)$$

where

$$g_{\text{LO}}(s) \stackrel{\text{def}}{=} \frac{-f'(s)}{f(s)} \quad (6)$$

### 3. DEFINITIONS OF RANKS

Let us define function  $S_1$  as follows:

$$S_1(v) = \begin{cases} 1 & v > 0 \\ 0 & v = 0 \\ -1 & v < 0 \end{cases} \quad (7)$$

Let  $\mathbf{R}_1[n]$  denote the normalized marginal rank vector of  $\mathbf{y}[n]$  based on  $\{\mathbf{y}[n], \dots, \mathbf{y}[n+L-1]\}$ , defined in [6],

$$\mathbf{R}_1[n] = \frac{1}{L-1} \sum_{i=n}^{n+L-1} S_1(\mathbf{y}[n] - \mathbf{y}[i]) \quad (8)$$

Here  $S_1$  is applied component-wise, and  $L$  is the number of samples chosen to compute the normalized marginal ranks of  $\mathbf{y}[n]$ .

The normalized marginal ranks defined in (8) are only for real-valued data. However, our second definition of ranks can be applied to both real-valued and complex-valued data. Let us first define:

$$S_2(v) = \begin{cases} 1 & v \geq 0 \\ 0 & v < 0 \end{cases} \quad (9)$$

and

$$\text{sign}(u) = \frac{u}{|u|} \quad (10)$$

Define the normalized signed-rank vector  $\mathbf{R}_2[n]$  of  $\mathbf{y}[n]$  based on  $\{\mathbf{y}[n], \dots, \mathbf{y}[n+L-1]\}$  to be:

$$\mathbf{R}_2[n] = \text{sign}(\mathbf{y}[n]) \left\{ \frac{1}{L} \sum_{i=n}^{n+L-1} S_2(|\mathbf{y}[n]| - |\mathbf{y}[i]|) \right\} \quad (11)$$

where  $S_2$  and  $\text{sign}(\cdot)$  are applied component-wise, and  $L$  is the number of samples chosen to compute the rank vector of  $\mathbf{y}[n]$ .

We note that the ranks defined in (8) are the normalized and centered version of the traditional ranks which are in the range  $\{1, \dots, L\}$ . For simplicity, let us assume we have a set of real values  $\{x_1, \dots, x_L\}$ . Assuming  $x_{j_1} < x_{j_2} < \dots < x_{j_L}$  with  $j_1, j_2, \dots, j_L \in \{1, 2, \dots, L\}$ , the traditional method ranks  $x_{j_1}, x_{j_2}, \dots, x_{j_L}$  with intergers  $1, 2, \dots, L$  respectively. Denote  $r_i$  with  $i = 1, \dots, L$  as the rank of  $x_i$  in the traditional definition; let  $R_{1i}$  denote the normalized marginal rank of  $x_i$  defined in this paper. The explicit definition of  $r_i$  is given as follows:

$$r_i = \sum_{j=1}^L S_2(x_i - x_j) \quad (12)$$

where  $S_2(\cdot)$  is defined in (9). It is easy to see that

$$R_{1i} = \frac{2r_i - (L+1)}{L-1} \quad (13)$$

with  $r_i \in \{1, 2, \dots, L\}$  and  $R_i \in \{-1, -\frac{L-3}{L-1}, \dots, \frac{L-3}{L-1}, 1\}$ .

Assuming that we have a set of numbers  $\{z_1, \dots, z_L\}$ , let  $z_i = a_i e^{j\theta_i}$  with  $i = 1, \dots, L$ , where  $a_i$  and  $\theta_i$  are the amplitude and phase of  $z_i$  respectively. Let us denote  $r_{ai}$  with  $i = 1, \dots, L$  as the rank of  $a_i$  in the traditional definition based on  $a_1, \dots, a_L$ ; let  $R_{2i}$  denote the normalized signed-rank of  $z_i$  based on  $\{z_1, \dots, z_L\}$ , as defined in (11). Thus we can see that

$$R_{2i} = \frac{1}{L} r_{ai} \text{sign}(z_i) \quad (14)$$

Therefore, in the above equation, the phase information of each  $z_i$  is retained in the term  $\text{sign}(z_i)$  along with the traditional rank representation for its modulus. The factor  $\frac{1}{L}$  is used to obtain a normalized signed rank representation.

#### 4. MODIFIED ALGORITHM USING RANKS

Our modified algorithm using ranks is

$$\begin{aligned} \mathbf{B}_{n+1} = & \mathbf{B}_n - \mathbf{B}_n \left\{ \lambda_n [\mathbf{y}[n]\mathbf{y}[n]^T - \mathbf{I}] \right. \\ & \left. + \mu_n [\mathbf{h}(\mathbf{R}_\mathbf{y}[n])\mathbf{R}_\mathbf{y}^T[n] - \mathbf{R}_\mathbf{y}[n]\mathbf{h}^T(\mathbf{R}_\mathbf{y}[n])] \right\} \end{aligned} \quad (15)$$

where  $\lambda_n$  and  $\mu_n$  are adaptation steps, and  $\mathbf{h}(\cdot)$  is a component-wise nonlinear odd rank weighting function;  $\mathbf{R}_\mathbf{y}[n]$  is the rank vector of  $\mathbf{y}[n]$  based on  $\{\mathbf{y}[n], \dots, \mathbf{y}[n+L-1]\}$ . If  $\mathbf{y}[n]$  is real-valued,  $\mathbf{R}_\mathbf{y}[n]$  could be the normalized marginal rank vector defined in (8) or the normalized signed-rank vector defined in (11). On the other hand, if  $\mathbf{y}[n]$  is complex-valued,  $\mathbf{R}_\mathbf{y}[n]$  will be normalized signed-rank vector for  $\mathbf{y}[n]$ .

If the components of  $\mathbf{y}[n]$  are mutually independent, then  $\mathbf{R}_\mathbf{y}[n]$  components for fixed  $n$  are also independent. Assuming  $R_{yi}[n]$  and  $h(R_{yj}[n])$  are the  $i$ th and the  $j$ th components of  $\mathbf{R}_\mathbf{y}[n]$  and  $\mathbf{h}(\mathbf{R}_\mathbf{y}[n])$ , then, by the independence of  $R_{yi}[n]$  and  $R_{yj}[n]$  with  $i \neq j$  and  $i, j \in \{1, \dots, m\}$ , we have  $\mathbf{E}(R_{yi}[n]h(R_{yj}[n])) = \mathbf{E}(R_{yi}[n])\mathbf{E}(h(R_{yj}[n]))$ .

Assuming  $\mathbf{R}_\mathbf{y}[n]$  is the normalized marginal rank vector, then each  $R_{yj}[n]$  with  $j = 1, \dots, m$  equals  $(L+1-2i)/(L-1)$ ,  $i = 1, \dots, L$  with probability  $\frac{1}{L}$ . It is easy to see in this case that

$$\begin{aligned} \mathbf{E}(R_{yj}[n]) &= \sum_{i=1}^L \frac{L+1-2i}{L-1} \cdot \frac{1}{L} \\ &= 0 \end{aligned} \quad (16)$$

Assume  $\mathbf{R}_\mathbf{y}[n]$  is the normalized signed-rank vector of  $\mathbf{y}[n]$  based on  $\{\mathbf{y}[n], \dots, \mathbf{y}[n+L-1]\}$ . Let  $y_j[n] = |y_j[n]|e^{j\theta_j}$  be the  $j$ -th component of  $\mathbf{y}[n]$ , where  $\theta_j$  is the phase of  $y_j[n]$ . If the real part and the imaginary part of  $y_j[n]$  are circularly symmetric in their joint pdf, the magnitude  $|y_j[n]|$  and the phase  $\theta_j$  of  $y_j[n]$  are independent and  $\mathbf{E}(e^{j\theta_j}) = 0$ . Denote  $r_{aj}$  as the traditional rank of  $|y_j[n]|$  based on  $\{|y_j[n]|, \dots, |y_j[n+L-1]|\}$ . It is easy to see that  $r_{aj}$  is also independent of  $\theta_j$ . Thus we have

$$\begin{aligned} \mathbf{E}(R_{yj}[n]) &= \frac{1}{L} \mathbf{E}(r_{aj} \cdot e^{j\theta_j}) \\ &= \frac{1}{L} \mathbf{E}(r_{aj})\mathbf{E}(e^{j\theta_j}) \\ &= 0 \end{aligned} \quad (17)$$

Therefore, in both cases where  $R_{yj}[n]$  is either the normalized marginal rank or the normalized signed rank,  $\mathbf{E}(R_{yi}[n]h(R_{yj}[n])) = 0$  with  $i \neq j$  and  $i, j \in \{1, \dots, m\}$ . By forcing this condition together with the whitening condition  $\mathbf{E}[\mathbf{y}[n]\mathbf{y}[n]^T] = \mathbf{I}$  in the algorithm, it is possible to drive the components of  $\mathbf{y}[n]$  to be independent.

The normalized marginal and signed ranks defined in (8) and (11) respectively can greatly increase the stability and robustness of the algorithm. Therefore, the modified algorithm in (15) does not need the normalization factors, at least in the second term. Simulations using this modified algorithm with no normalization at all also show stable and robust performance.

The term  $(\mathbf{h}(\mathbf{R}_\mathbf{y}[n])\mathbf{R}_\mathbf{y}^T[n] - \mathbf{R}_\mathbf{y}[n]\mathbf{h}^T(\mathbf{R}_\mathbf{y}[n]))$  in (15) can drive the off-diagonal elements of  $\mathbf{C}_n$  defined in (3) to all zeros. However, the convergence rate of these off-diagonal elements may be slower using ranks. Thus, without comprising stability in (15),  $\mu_n$  may be chosen to be greater than  $\lambda_n$  to increase the convergence rate of these off-diagonal elements compared to that of the diagonal elements.

It is easy to see that the normalized signed ranks defined in (11) are inside or on the unit circle. Thus the amplitudes of the original data have been largely compressed to avoid the "blow-up" of BSS algorithms. However, the phase information has been retained by our definition in (11).

#### 5. NONLINEAR RANK WEIGHTING FUNCTIONS

We can choose the rank weighting function  $h(R) = g_{LO}(R)$  with  $g_{LO}(\cdot)$  defined in (6) and the ranks  $R$  defined in this paper. A more detailed discussion on the choice of the rank weighting functions is given in [5].

Consider a unit-variance generalized Gaussian source pdf of the form  $f(s) = C(k)e^{-c(k)|s|^k}$  with  $k > 0$ , where

$c(k)$  and  $C(k)$  are two constants when  $k$  is fixed. Thus the optimum  $g$  function given in (5) will be  $g_{LO} = \alpha(k)|s|^{k-1}\text{sign}(s)$  where  $\alpha(k)$  is a function of  $k$ . Since  $\alpha(k)$  is a constant for any fixed  $k$ , it can be accommodated by the adaptation step size  $\mu_n$  defined in (15). Therefore, we choose  $h(R) = |R|^{k-1}\text{sign}(R)$  in this case. However, our simulations show that multiple choices of weighting functions are applicable and all give generally good performance. More generally, considering the class of generalized Gaussian signals, we may choose  $h(R) = \text{sign}(R)|R|^m$  with  $m \geq k-1$  for  $k > 2$ , and  $h(R) = \text{sign}(R)|R|^m$  with  $m \leq k-1$  for  $k < 2$ .

Let us consider the case when the received signals are complex-valued data. In [2], the nonlinear  $g$  functions are restricted to be of the form  $g(y) = y \cdot l(|y|^2)$  if  $y$  is complex, where  $l(\cdot)$  is a real-valued function. Since the normalized signed ranks for complex numbers are still complex-valued, we propose the nonlinear rank weighting functions in the complex case to be  $h(R) = g(R)$ . Our simulation results in the next section show good performance with our choice of nonlinear rank weighting functions.

## 6. SIMULATIONS

We present here representative simulation results for the algorithms discussed in this paper. Performance comparisons between the EASI algorithms and our modified algorithm are given in the following examples. Both examples have two sources and two receivers. The mixing matrix  $\mathbf{A}$  is randomly generated for both examples. In each example, the original/normalized EASI algorithm and the modified algorithm are run for the same number of iterations and for the same set of signals. Fig. 3 and Fig. 5 show the elements of the  $\mathbf{C}_n$  defined in (3) as a function of the number of iterations.

### Example 1

In our simulation for Fig. 3, the two sources are two 16QAM sequences. The original source symbols are contaminated by heavy-tailed non-Gaussian noise. Vectorizing the samples taken at the receivers, we have

$$\mathbf{x}[n] = \mathbf{A}\mathbf{p}[n]$$

where  $\mathbf{x}[n]$  is the vector of all the samples taken at time index  $n$  at the receivers, and

$$\mathbf{p}[n] = \mathbf{s}[n] + \mathbf{v}[n]$$

where  $\mathbf{s}[n]$  is the vector of all the original source symbols at time index  $n$ , and  $\mathbf{v}[n]$  is the vector of complex additive non-Gaussian noise at time index  $n$ .

Let us denote  $v_i[n] = a_i + jb_i$  with  $i = 1, 2$  as the  $i$ -th component of vector  $\mathbf{v}[n]$ , where  $a_i$  and  $b_i$  are the real part and the imaginary part of  $v_i[n]$  respectively. In this example,  $a_i$  and  $b_i$  are independent and identically distributed with the pdf  $0.9\mathcal{N}(0, 7/9) + 0.1\mathcal{N}(0, 3)$ , where  $\mathcal{N}(\mu, \sigma^2)$  is a Gaussian pdf with mean  $\mu$  and variance  $\sigma^2$ . Fig. 2 shows the noise pdf's around each of 16QAM symbols.

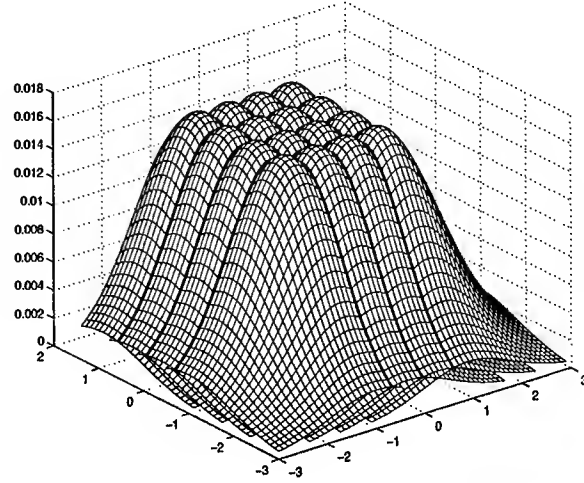


Figure 2: The non-Gaussian noise pdf's for 16 QAM symbols

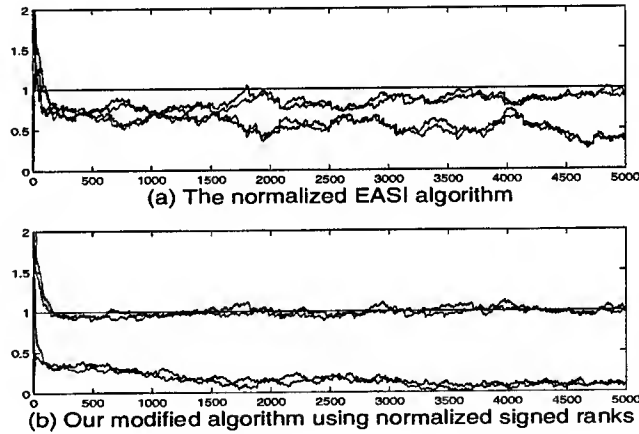


Figure 3: **Example 1:** Example Performance for 16QAM symbols in presence of heavy-tailed non-Gaussian noise

The channel characteristic between  $\mathbf{p}[n]$  and  $\mathbf{x}[n]$  is defined by the mixing matrix  $\mathbf{A}$  with

$$\mathbf{A} = \begin{pmatrix} -0.0783 - 0.0118i & 2.3093 + 0.0559i \\ 0.8892 + 0.9131i & 0.5246 - 1.1071i \end{pmatrix}$$

Fig. 3 shows the results of running the normalized

EASI algorithm and the modified algorithm. The adaptation step is chosen as  $\lambda_n = 0.005$  for the normalized EASI algorithm, while  $\lambda_n = 0.005$  and  $\mu_n = 0.035$  for the modified algorithm in (15) using normalized signed ranks. Simulations for example 1 have shown that our modified algorithm gives generally good results. However, the original EASI algorithm seems to blow up in our simulations. We also find that the normalized EASI algorithm exhibits poor performance.

**Example 2** In our simulation for Fig. 5, two 16QAM source sequences are transmitting through a communication channel on different subcarrier frequencies. There exists cochannel interference due to the overlap of the frequency bands. The channel model is shown in Fig. 4 with the overlap ratio  $k/2W = 13\%$ . Samples are taken at each receiver with symbol rate  $1/T$  without ISI. Heavy-tailed non-Gaussian noise is added at the receiver with SNR=10dB.

Fig. 5 shows the results of running the original EASI algorithm and the modified algorithm. The adaptation step is chosen as  $\lambda_n = 0.005$  for the original EASI algorithm, while  $\lambda_n = 0.005$  and  $\mu_n = 0.02$  for the modified algorithm using normalized signed ranks. We can see our modified algorithm shows better performance than the original EASI algorithm.

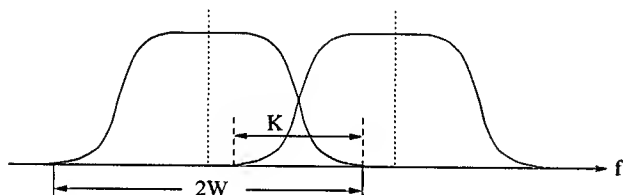


Figure 4: FDMA Channel Model with Cochannel Overlap

## 7. CONCLUSION

In this paper, we have proposed a new approach using ranks to improve the robustness of the EASI algorithms. Choosing different ranking methods, our modified algorithm can be applied to either real-valued data or complex-valued data. We also give some guidelines for designing the nonlinear rank weighting functions used in our modified algorithm. We have shown that our approach is more robust in estimating the mixing channel for source separation as compared to the original algorithm. More studies need to be done on the general optimum rank weighting functions for both real-valued data and complex-valued data.

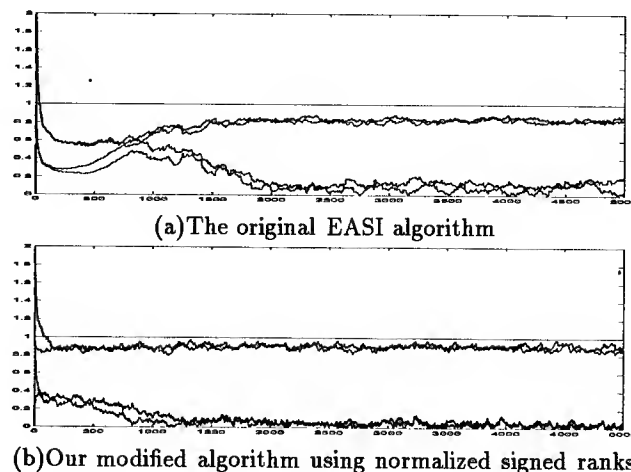


Figure 5: Performance for 16QAM symbols in FDMA with cochannel interference in the presence of non-Gaussian noise

## 8. REFERENCES

- [1] P. Comon, "Independent component analysis. A new concept?" *Signal Processing*, vol. 36, pp. 287-314, Apr. 1994.
- [2] J.F. Cardoso and B. Laheld, "Equivariant adaptive source separation," *IEEE Trans. on Signal Processing*, vol. 44, pp. 3017-3030, Dec. 1996.
- [3] J.F. Cardoso, "Blind source separation: Statistical principles," *Proc. IEEE*, 86, pp. 2009-2025, Oct. 1996.
- [4] S. Kassam and Y. Zhang, "Some results on a BSS algorithm under non-standard conditions," *Proc. Conf. on Information Sciences and Systems*, Mar. 1999.
- [5] Y. Zhang and S. Kassam, "Robust source separation based on ranks," *Proc. Conf. on Information Sciences and Systems*, Mar. 2000.
- [6] S. Visuri, V. Koivunen and H. Oja, "Sign and rank covariance matrices," *Journal of Statistical Inference and Planning*, to appear, 1999.
- [7] J. Hájek and Z. Šidák, "Theory of rank tests," *Academic Press*, 1967.

# SEMI-BLIND MAXIMUM LIKELIHOOD SEPARATION OF LINEAR CONVOLUTIVE MIXTURES

João Xavier and Victor Barroso

Instituto Superior Técnico – Instituto de Sistemas e Robótica  
Av. Rovisco Pais, 1049-001 Lisboa Codex, Portugal  
{jxavier,vab}@isr.ist.utl.pt

## ABSTRACT

We address the problem of separating a linear convolutive mixture of 2nd order white sources, given some side information about the transmitted messages. The proposed technique exploits the special structure of the observed data matrix, after channel whitening: it is the product of an orthogonal and generalized Toeplitz matrices in additive Gaussian noise. We implement the joint maximum likelihood (ML) estimator of both the orthogonal mixing matrix and the user signals, subject to the known algebraic and temporal constraints. Preliminary computer simulations assess the promising performance of the proposed method.

## 1. INTRODUCTION

Linear convolutive mixtures of sources occur in many scenarios of interest. For example, in space division multiple access (SDMA) wireless networks for mobile communications, the signal observed at the base station receiver is a weighted linear superposition of the emitted user signals plus echos [1, 2, 3, 4, 5, 6]. This is due to the multipath propagation effect (intersymbol interference phenomenon), and the fact that several sources share the same carrier frequency (co-channel sources) for bandwidth efficiency (SDMA concept). The SDMA receiver must resolve the observed mixture, and recover the transmitted signals. In this paper, we introduce a maximum-likelihood (ML) technique to resolve linear convolutive mixtures. The proposed technique is semi-blind because certain fragments of the emitted messages are assumed known. The side information is necessary because we do not restrict ourselves to finite-alphabet sources. As a consequence, in the absence of the side information, the factorization in the data model would not uniquely determine both the channel and the user signals. The paper is organized as follows. In section 2, we establish the data model and define the problem formulation. In section 3, we develop the iterative maximum-likelihood (IML) algorithm which estimates both the orthogonal mixing matrix and the user signals. Certain algorithmic issues are briefly discussed. In section 4, we present computer simulation results assessing the performance of the proposed ML technique. Section 5 concludes our paper.

**Notation.** Matrices (capital) and vectors are in boldface type.  $\mathbb{R}^{n \times m}$  is the set of  $n \times m$  matrices with real entries.  $(\cdot)^T$ ,  $(\cdot)^+$ ,  $\otimes$ ,  $\text{tr}\{\cdot\}$ , and  $\text{vec}(\cdot)$  stand for the transpose, Moore Penrose pseudo inverse, Kronecker product,

the trace, and the vectorization operator, respectively. For a matrix  $\mathbf{A} \in \mathbb{R}^{n \times m}$ ,  $\|\mathbf{A}\| = \sqrt{\text{tr}\{\mathbf{A}^T \mathbf{A}\}}$  denotes its Frobenius norm.  $\mathbf{I}_n$  and  $\mathbf{0}_{n \times m}$  represent the  $n \times n$  identity matrix and the all-zero  $n \times m$  matrix, respectively (when the dimensions are clear from the context, the subscripts are dropped). For a vector  $\boldsymbol{\theta} = [\theta_1 \theta_2 \cdots \theta_l]^T \in \mathbb{R}^l$ ,  $\mathcal{T}_{n \times m}(\boldsymbol{\theta})$  denotes the  $n \times m$  Toeplitz matrix generated by  $\boldsymbol{\theta}$ , i.e.,

$$\mathcal{T}_{n \times m}(\boldsymbol{\theta}) = \begin{bmatrix} \theta_n & \theta_{n+1} & \theta_{n+2} & \cdots & \theta_l \\ \theta_{n-1} & \theta_n & \cdots & \cdots & \theta_{l-1} \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \theta_1 & \theta_2 & \cdots & \cdots & \theta_{l-n+1} \end{bmatrix},$$

where  $m + n - 1 = l$ .  $\mathbf{x} \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{C})$  means that the random vector  $\mathbf{x}$  is Gaussian distributed with mean  $\boldsymbol{\mu}$  and covariance matrix  $\mathbf{C}$ .

## 2. PROBLEM STATEMENT

Consider  $P$  co-channel sources, observed through a convolutive finite-impulse response (FIR) multichannel system:

$$\mathbf{x}(k) = \sum_{p=1}^P \sum_{l=0}^{L-1} \mathbf{h}_p(l) \theta_p(k-l) + \mathbf{w}(k); \quad (1)$$

$\mathbf{x}(k) \in \mathbb{R}^N$  denotes the observations,  $\mathbf{h}_p(l) \in \mathbb{R}^N$  for  $l = 0, 1, \dots, L-1$ , is the FIR of the  $p$ th multichannel filter (for simplicity, all FIR multichannels have the same length  $L$ ),  $\theta_p(k) \in \mathbb{R}$  is the scalar signal transmitted by the  $p$ th user, and  $\mathbf{w}(k) \in \mathbb{R}^N$  represents additive noise. Rewrite (1) as

$$\mathbf{x}(k) = \underbrace{[\mathbf{H}_1 \mathbf{H}_2 \cdots \mathbf{H}_P]}_{\mathbf{H}} \underbrace{\begin{bmatrix} \theta_1(k) \\ \theta_2(k) \\ \vdots \\ \theta_P(k) \end{bmatrix}}_{\boldsymbol{\theta}(k)} + \mathbf{w}(k), \quad (2)$$

where,  $\mathbf{H}_p = [\mathbf{h}_p(0) \mathbf{h}_p(1) \cdots \mathbf{h}_p(L-1)] : N \times L$ , and  $\boldsymbol{\theta}_p(k) = [\theta_p(k) \theta_p(k-1) \cdots \theta_p(k-L+1)]^T : L \times 1$ .  $\mathbf{H} : N \times LP$  is the unknown convolutive mixing matrix. Given a finite set of observations  $\mathcal{X} = \{\mathbf{x}(k) : k = 1, 2, \dots, K\}$ , we aim at estimating the user signals  $\theta_p(k)$ . The following assumptions are assumed to hold throughout the paper. (A1)  $P$  is known, and  $\mathbf{H}$  is full column-rank. (A2) The sources

are uncorrelated, zero-mean and white up to 2nd order, *i.e.*,  $r_{p,q}(k,l) = \mathbb{E}\{\theta_p(k)\theta_q(l)\} = \delta(p-q)\delta(k-l)$ , where  $\delta(\cdot)$  is the Kronecker delta. No statistical information about the sources, beyond the 2nd order, is assumed known. (A3)  $w(k)$  denotes zero-mean spatio-temporal white Gaussian noise with known power  $\sigma^2$ . As it is well known, even for noiseless scenarios, the factorization  $H\theta(k)$  in the data model (2), is not unique:  $H$  can be solved only up to a residual  $P \times P$  instantaneous mixing matrix [4, 6]. To guarantee uniqueness of the factorization in (2), we assume that (A4) for each source  $p$ , a certain fragment  $\mathcal{F}_p(i_p, j_p) = \{\theta_p(i_p), \theta_p(i_p+1), \dots, \theta_p(j_p)\}$  is known. Here,  $i_p$  and  $j_p$  denote the beginning and end of the known excerpt, respectively. It is not required that  $i_p = i_q$  for  $q \neq p$  (nor  $j_p = j_q$ ), *i.e.*, we do not assume synchronization among the sources (difficult to ensure in practice), but only vis-a-vis each source and the base station receiver.

### 3. THE IML ALGORITHM

**Channel whitening.** We work with whitened data samples. Consider the eigenvalue-decomposition (EVD),

$$\mathbf{R}_x = \mathbb{E}\{\mathbf{x}(k)\mathbf{x}(k)^T\} = \underbrace{\mathbf{U}\mathbf{\Sigma}\mathbf{U}^T}_{\mathbf{H}\mathbf{H}^T} + \sigma^2 \mathbf{I}_N, \quad (3)$$

where  $\mathbf{U} = [\mathbf{U}_1 \mathbf{U}_2] \in \mathcal{O}(N)$ ,  $\mathbf{U}_1 : N \times LP$ , and  $\mathbf{\Sigma} = \text{diag}(\mathbf{\Sigma}_1, \mathbf{0})$ ; here, and for future reference,

$$\mathcal{O}(N) = \{\mathbf{W} \in \mathbb{R}^{N \times N} : \mathbf{W}^T \mathbf{W} = \mathbf{I}_N\},$$

denotes the group of  $N \times N$  orthogonal matrices. It is readily seen that

$$\mathbf{P} \equiv \mathbf{U}_1 \mathbf{\Sigma}_1^{1/2} = \mathbf{H}\mathbf{Q}^T, \quad (4)$$

where  $\mathbf{Q}$  denotes an (unknown) residual orthogonal matrix. The whitened data samples are given by

$$\mathbf{y}(k) = \mathbf{P}^+ \mathbf{x}(k) = \mathbf{Q}\boldsymbol{\theta}(k) + \mathbf{n}(k), \quad k = 1, \dots, K, \quad (5)$$

where  $\mathbf{n}(k) = \mathbf{P}^+ w(k) \sim \mathcal{N}(\mathbf{0}, \mathbf{C})$ ,  $\mathbf{C} = \sigma^2 \mathbf{\Sigma}_1^{-1}$ . The principal advantage of the whitened data model is that the channel matrix  $\mathbf{Q} : LP \times LP$  has less parameters to estimate than the corresponding channel matrix  $\mathbf{H} : N \times LP$  in (2); also, it is more structured (orthogonal). For the  $p$ th source, we collect all the unknowns in (5) in  $\boldsymbol{\theta}_p \equiv [\theta_p(2-L) \dots \theta_p(K)]^T$ . Further, the resulting  $P$  vectors are stacked in  $\boldsymbol{\theta} \equiv [\boldsymbol{\theta}_1^T \dots \boldsymbol{\theta}_P^T]^T$ . Also, we express the knowledge of the fragment  $\mathcal{F}_p(i_p, j_p)$  in matrix terms as  $\mathbf{E}_p^T \boldsymbol{\theta}_p = \boldsymbol{\eta}_p$ , where  $\mathbf{E}_p$  selects the appropriate entries of  $\boldsymbol{\theta}_p$ , and  $\boldsymbol{\eta}_p$  contains the *a priori* known values, *i.e.*,  $\mathbf{E}_p : (K+L-1) \times (j_p - i_p + 1)$  contains the columns  $L-1+i_p$  to  $L-1+j_p$  of  $\mathbf{I}_{K+L-1}$ , and  $\boldsymbol{\eta}_p = [\theta_p(i_p) \dots \theta_p(j_p)]^T$ . In terms of the overall vector of unknowns  $\boldsymbol{\theta}$ , we have  $\mathbf{E}^T \boldsymbol{\theta} = \boldsymbol{\eta}$ , where  $\mathbf{E} = \text{diag}(\mathbf{E}_1, \dots, \mathbf{E}_P)$  and  $\boldsymbol{\eta} = [\boldsymbol{\eta}_1^T \dots \boldsymbol{\eta}_P^T]^T$ .

**IML Algorithm.** Notice that, by assumption (A2), the receiver ignores the statistical description of the sources beyond the 2nd order. Thus,  $\boldsymbol{\theta}$  is viewed, in the sequel, as a deterministic vector of unknowns. We aim at finding the

joint ML estimator of  $(\mathbf{Q}, \boldsymbol{\theta})$ , subject to the known constraints, *i.e.*,

$$(\hat{\mathbf{Q}}, \hat{\boldsymbol{\theta}}) = \arg \max_{\mathbf{Q} \in \mathcal{O}(LP), \mathbf{E}^T \boldsymbol{\theta} = \boldsymbol{\eta}} l(\mathbf{y}(1), \dots, \mathbf{y}(K) | \mathbf{Q}, \boldsymbol{\eta}),$$

where  $l(\cdot | \mathbf{Q}, \boldsymbol{\eta})$  stands for the (conditioned) likelihood of the whitened observations. After some algebra, the optimization task at hand can be formulated as

$$(\hat{\mathbf{Q}}_{ML}, \hat{\boldsymbol{\theta}}_{ML}) = \arg \min_{\mathbf{Q} \in \mathcal{O}(LP), \mathbf{E}^T \boldsymbol{\theta} = \boldsymbol{\eta}} \phi(\mathbf{Q}, \boldsymbol{\theta}), \quad (6)$$

where

$$\phi(\mathbf{Q}, \boldsymbol{\theta}) = \frac{1}{K} \|\mathbf{Y} - \mathbf{Q}\boldsymbol{\mathcal{T}}(\boldsymbol{\theta})\|_{\mathbf{C}^{-1}}^2. \quad (7)$$

$\mathbf{Y} = [\mathbf{y}(1) \dots \mathbf{y}(K)]$  denotes the observed whitened data matrix,  $\boldsymbol{\mathcal{T}}(\boldsymbol{\theta})$  is the *generalized* Toeplitz matrix generated by  $\boldsymbol{\theta}$ ,  $\boldsymbol{\mathcal{T}}(\boldsymbol{\theta}) = [\boldsymbol{\mathcal{T}}_{L_1 \times K}(\boldsymbol{\theta}_1)^T \dots \boldsymbol{\mathcal{T}}_{L_P \times K}(\boldsymbol{\theta}_P)^T]^T$ , and  $\|\mathbf{Z}\|_{\mathbf{C}^{-1}}^2 = \text{tr}\{\mathbf{Z}^T \mathbf{C}^{-1} \mathbf{Z}\}$ , for arbitrary  $\mathbf{Z} \in \mathbb{R}^{L \times K}$ . We propose an iterative procedure which minimizes  $\phi(\mathbf{Q}, \boldsymbol{\theta})$  with respect to  $\mathbf{Q}$  and  $\boldsymbol{\theta}$ , cyclically. Table 1 lists the resulting iterative maximum-likelihood (IML) algorithm. Thus,

Let  $\mathbf{Q}^{(0)} \in \mathcal{O}(LP)$   
for  $n = 1, 2, \dots$   
i)  $\boldsymbol{\theta}^{(n)} = \arg \min_{\mathbf{E}^T \boldsymbol{\theta} = \boldsymbol{\eta}} \phi(\mathbf{Q}^{(n-1)}, \boldsymbol{\theta})$   
ii)  $\mathbf{Q}^{(n)} = \arg \min_{\mathbf{Q} \in \mathcal{O}(LP)} \phi(\mathbf{Q}, \boldsymbol{\theta}^{(n)})$   
until  $\mathbf{Q}^{(n)} - \mathbf{Q}^{(n-1)} = \mathbf{0}$

Table 1: IML Algorithm

the IML algorithm is a cyclic coordinate descent method, and only locally convergent.

**Solving for  $\boldsymbol{\theta}^{(n)}$ .** Write  $\boldsymbol{\mathcal{T}}(\boldsymbol{\theta}) = \sum_{p=1}^P \sum_{l=2-L}^K \mathbf{B}_p(l) \theta_p(l)$ , the matrices  $\mathbf{B}_p(\cdot) : LP \times K$  being implicitly defined; thus,  $\text{vec}(\boldsymbol{\mathcal{T}}(\boldsymbol{\theta})) = \mathbf{B}\boldsymbol{\theta}$ ,  $\mathbf{B} = [\mathbf{b}_1(2-L) \dots \mathbf{b}_P(K)]$ ,  $\mathbf{b}_p(l) = \text{vec}(\mathbf{B}_p(l))$ , and

$$\phi(\mathbf{Q}, \boldsymbol{\theta}) = \frac{1}{K} \|\mathbf{y} - (\mathbf{I}_K \otimes \mathbf{Q}) \mathbf{B}\boldsymbol{\theta}\|_{\mathbf{I}_K \otimes \mathbf{C}^{-1}}^2, \quad (8)$$

with  $\mathbf{y} = \text{vec}(\mathbf{Y})$ . To minimize (8) with respect to  $\boldsymbol{\theta}$  ( $\mathbf{E}^T \boldsymbol{\theta} = \boldsymbol{\eta}$ ), write  $\boldsymbol{\theta} = \mathbf{E}\boldsymbol{\eta} + \mathbf{F}\boldsymbol{\lambda}$ ; here,  $\boldsymbol{\theta}_0 = \mathbf{E}\boldsymbol{\eta}$  and  $\mathbf{F}\boldsymbol{\lambda}$  represent the known and unknown parts, respectively. Notice that  $[\mathbf{E}\mathbf{F}]$  is an identity matrix (up to permutation of columns). Solving for  $\boldsymbol{\lambda}$  in (8), gives  $\boldsymbol{\lambda} = (\mathbf{F}^T \mathbf{A} \mathbf{F})^{-1} \mathbf{F}^T \mathbf{d}$ , where  $\mathbf{A} = \mathbf{B}^T (\mathbf{I}_K \otimes \mathbf{Q}^T \mathbf{C}^{-1} \mathbf{Q}) \mathbf{B}$ , and the  $i$ th entry of the vector  $\mathbf{d}$ ,  $d_i = \text{tr}\{\mathbf{Q}^T \mathbf{C}^{-1} \Delta_{\mathbf{Y}} \mathbf{B}_i^T\}$ , where  $\Delta_{\mathbf{Y}} = \mathbf{Y} - \mathbf{Q}\boldsymbol{\mathcal{T}}(\boldsymbol{\theta}_0)$ ; the matrices  $\mathbf{B}_i$  are defined by

$$\text{vec}(\mathbf{B}_i) = \mathbf{b}_i, \quad \mathbf{B} = [\mathbf{b}_1 \dots \mathbf{b}_{P(K-1)+LP}]. \quad (9)$$

**Solving for  $\mathbf{Q}^{(n)}$ .** After trivial algebraic manipulations, we face the problem of minimizing

$$\varphi(\mathbf{Q}) = \text{tr}\{\mathbf{Q}^T \mathbf{C}^{-1} \mathbf{Q} \hat{\mathbf{R}}_{\boldsymbol{\theta}}\} - 2 \text{tr}\{\mathbf{Q}^T \mathbf{C}^{-1} \hat{\mathbf{R}}_{\mathbf{y}\boldsymbol{\theta}}\}, \quad (10)$$

where  $\hat{R}_\theta = \frac{1}{K} \mathcal{T}(\theta) \mathcal{T}(\theta)^T$  and  $\hat{R}_{y\theta} = \frac{1}{K} Y \mathcal{T}(\theta)^T$ , subject to  $Q \in \mathcal{O}(LP)$ , the group of  $LP \times LP$  orthogonal matrices. It is well-known that  $\mathcal{O}(LP) \subset \mathbb{R}^{LP \times LP}$  is a differentiable manifold (of dimension  $LP(LP-1)/2$ ) [7]. In order to exploit the curvature of this constraint surface, we employ a geodesic descent algorithm (the generalization of the traditional steepest gradient method in flat spaces) [9]. Table 2 describes the idealized geodesic descent algorithm. In words, the algorithm proceeds by solving successive one-

1. Choose  $Q^{(0)} \in \mathcal{O}(LP)$
2. for  $m = 1, 2, \dots$ 
  - a) Let  $D$  denote the projection of  $-\nabla\varphi(Q^{(m-1)})$  onto the tangent space of  $\mathcal{O}(LP)$  at  $Q^{(m-1)}$
  - b) Let  $Q(t) \in \mathcal{O}(LP)$ ,  $t \geq 0$ , denote the geodesic emanating from  $Q(0) = Q^{(m-1)}$  in the direction  $\dot{Q}(0) = D$
  - c) Minimize  $\varphi(Q(t))$  with respect to  $t \geq 0$ , to obtain  $t_{\min}$  (a global minimizer); set  $Q^{(m)} = Q(t_{\min})$
3. until  $Q^{(m)} - Q^{(m-1)} = 0$

Table 2: Geodesic descent algorithm

parameter geodesic minimization problems (as line search methods do in flat spaces [9]). We now focus on the details of each of the sub-steps a), b) and c) of the geodesic descent algorithm.

**Sub-step a).** To simplify notation, and for future reference, let  $Q_0 = Q^{(m-1)}$ . By applying standard calculus rules, it can be seen that, the gradient (in matrix format) of  $\varphi(\cdot)$  evaluated at  $Q_0$ , is given by

$$\nabla\varphi(Q_0) = 2C^{-1} (Q_0 \hat{R}_\theta - \hat{R}_{y\theta}).$$

On the other hand, the tangent space to the manifold  $\mathcal{O}(LP)$  at the point  $Q_0 \in \mathcal{O}(LP)$  is given by

$$\mathcal{T}_{\mathcal{O}(LP)}(Q_0) = \{Q_0 K : K \in \mathcal{K}(LP)\},$$

see [7]; here  $\mathcal{K}(L) = \{K \in \mathbb{R}^{LP \times LP} : K = -K^T\}$  denotes the linear subspace of skew-symmetric matrices in  $\mathbb{R}^{LP \times LP}$ . Thus,  $D$ , the orthogonal projection of  $\tilde{\nabla} \equiv -\nabla\varphi(Q_0)$  onto the tangent space of  $\mathcal{O}(LP)$  at  $Q_0$  is equal to  $D = Q_0 K_0$ , where  $K_0$  can be found as

$$\begin{aligned} K_0 &= \arg \min_{K \in \mathcal{K}(LP)} \|Q_0 K - \tilde{\nabla}\| \\ &= \arg \min_{K \in \mathcal{K}(LP)} \|K - Q_0^T \tilde{\nabla}\| \\ &= \frac{1}{2} (Q_0^T \tilde{\nabla} - \tilde{\nabla}^T Q_0), \end{aligned}$$

the skew-symmetric component of the matrix  $Q_0^T \tilde{\nabla}$ .

**Sub-step b).** We must find a geodesic  $Q(t)$ ,  $t \geq 0$ , subject to the initial conditions:  $Q(0) = Q_0$  and  $\dot{Q}(0) = D =$

$Q_0 K_0$ . In order to stay in the constraint surface  $\mathcal{O}(LP)$ , the curve  $Q(t)$  must satisfy:

$$\dot{Q}(t) = Q(t) K(t), \quad (11)$$

where  $K(t) \in \mathcal{K}(LP)$ . Also, it can be shown (see [9]) that, for constant speed geodesics, the acceleration vector is orthogonal to the manifold. Thus:

$$\ddot{Q}(t) = Q(t) S(t), \quad (12)$$

where  $S(t) \in \mathcal{S}(LP) = \{S \in \mathbb{R}^{LP \times LP} : S = S^T\}$ , the linear subspace of symmetric matrices in  $\mathbb{R}^{LP \times LP}$  (remark that  $\mathcal{S}(LP)$  is orthogonal to  $\mathcal{K}(LP)$ ). Using the representation (11) in (12), gives

$$Q(t) K(t)^2 + Q(t) \dot{K}(t) = Q(t) S(t).$$

Eliminating  $Q(t)$  and re-arranging terms, we have

$$\dot{K}(t) = S(t) - K(t)^2,$$

i.e.,  $\dot{K}(t) \in \mathcal{S}(LP)$  (notice that  $K(t)^2 \in \mathcal{S}(LP)$ ). On the other hand,  $\dot{K}(t) \in \mathcal{K}(LP)$ , since  $K(t) \in \mathcal{K}(LP)$ . Thus,  $\dot{K}(t) = 0$ , and  $K(t) = K(0)$  is a constant matrix; in fact,  $K(t) = K_0$ , by the restriction on  $\dot{Q}(0)$ . Now, by (11) and the condition on  $Q(0)$ , we find  $Q(t) = Q_0 e^{K_0 t}$ .

**Sub-step c).** We have to minimize  $\gamma(t) \equiv \varphi(Q(t))$  over  $t \geq 0$ . Here, instead of performing an exact minimization, we propose to locate the first (perhaps only local) minimizer of  $\gamma(t)$ . This inaccurate (sub-optimal) scheme (often used in practice) permits to alleviate the computational burden and does not impair convergence of the overall algorithm (the outer minimization loop), if a sufficient degree of descent in  $\gamma(t)$  is achieved. We propose to locate (the first) point  $t_0 > 0$  such that  $\dot{\gamma}(t_0) = 0$ , by the bisection method. Notice that  $\dot{\gamma}(a^{(0)}) < 0$ ,  $a^{(0)} = 0$ . Let  $\delta > 0$  be given, and  $l$  the lowest integer such that  $\dot{\gamma}(l\delta) > 0$ . Set  $b^{(0)} = l\delta$ . We have the following procedure.

1. for  $m = 1, 2, \dots$

$$\text{Let } c = \frac{1}{2}(a^{(m-1)} + b^{(m-1)})$$

if  $\dot{\gamma}(c) = 0$  stop

if  $\dot{\gamma}(c) < 0$  then

$$a^{(m)} = c, \quad b^{(m)} = b^{(m-1)}$$

else

$$a^{(m)} = a^{(m-1)}, \quad b^{(m)} = c$$

2. until  $|b^{(m)} - a^{(m)}| < \epsilon$

3. Set  $t_0 = \frac{1}{2}(a^{(m)} + b^{(m)})$

Using standard calculus rules and the fact that  $K_0$  commutes with  $e^{K_0 t}$ , we have

$$\dot{\gamma}(t) = \text{tr} \left\{ e^{-K_0 t} C_0 e^{K_0 t} S_0 \right\} + 2 \text{tr} \left\{ e^{-K_0 t} C_0 Q_0^T \hat{R}_{y\theta} K_0 \right\},$$

where  $C_0 = Q_0^T C^{-1} Q_0$ , and  $S_0 = K_0 \hat{R}_\theta - \hat{R}_{y\theta} K_0$ . The expression for  $\dot{\gamma}(t)$  can also be easily found, and used to

check if the obtained  $t_0$  is, in fact, a local minimizer (if not, the algorithm must be restarted with a smaller  $\delta$ , say,  $\delta/2$ ).

**Initialization** The geodesic descent algorithm is only locally convergent. A (possible first) initialization is given by

$$\mathbf{Q}^{(0)} = \Pi_{\mathcal{O}(LP)} \left\{ \mathbf{C}^{-1} \hat{\mathbf{R}}_{\mathbf{y}\theta} \right\}. \quad (13)$$

Here,  $\Pi_{\mathcal{O}(LP)}(\mathbf{Z})$  is the (nonlinear) projection of the matrix  $\mathbf{Z} \in \mathbb{R}^{LP \times LP}$  onto the orthogonal group  $\mathcal{O}(LP)$ . It is computed as follows: let  $\mathbf{Z} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^T$  denote a singular-value decomposition (SVD) of  $\mathbf{Z}$ ; then,  $\Pi_{\mathcal{O}(LP)}(\mathbf{Z}) = \mathbf{U}\mathbf{V}^T$ . This remark is based on the fact that, near the global minimum,

$$\hat{\mathbf{R}}_{\theta} = \frac{1}{K} \sum_{k=1}^K \theta(k) \theta(k)^T \simeq \mathbf{R}_{\theta} = \mathbb{E} \left\{ \theta(k) \theta(k)^T \right\} = \mathbf{I}_{LP}.$$

Thus, the first term in (10) reduces to a constant, and  $\varphi(\mathbf{Q}) \simeq -2 \text{tr} \left\{ \mathbf{Q}^T \mathbf{C}^{-1} \hat{\mathbf{R}}_{\mathbf{y}\theta} \right\}$ , which is minimized by (13) [8].

**Binary sources.** For noiseless data samples and binary sources, i.e.,  $\theta_p(k) = \pm 1$ , the factorization implicit in (2) is essentially unique: the sources are determined up to a permutation and a sign ambiguity; see [1, 4]. Thus, prior knowledge of data fragments can be discarded. To take into account the finite-alphabet property of the sources, we can introduce in the IML algorithm the extra step i)a):  $\theta^{(n)} \leftarrow \text{sign} \left\{ \theta^{(n)} \right\}$ , i.e., the entries of  $\theta^{(n)}$  are projected onto the binary alphabet  $\mathcal{A} = \{\pm 1\}$ . This modification makes the IML algorithm more sensible to initialization, as monotone convergence of  $\phi$  can no longer be assured. Notice that the (optimal) approach of performing step i) with the entries of  $\theta$  restricted to  $\mathcal{A}$  is, from the computational viewpoint, considerably complex.

#### 4. EXPERIMENTAL RESULTS

To evaluate the performance of the IML algorithm, we conducted some computer simulations. We considered  $P = 2$  binary users, and a channel matrix  $\mathbf{H} : N \times LP$ , where  $N = 10$  and  $L = 3$ . The entries of  $\mathbf{H}$  were randomly generated (independent samples of a zero-mean Gaussian random variable with unit variance). Since the sources are binary, no symbol is assumed known *a priori*, and the IML algorithm is runned with the step i)a) (adaptation for BP-SK sources) discussed in section 3. The signal-to-noise ratio (SNR) is defined as  $\text{SNR} = \|\mathbf{H}\|^2 / (N\sigma^2)$ . The SNR was varied between  $\text{SNR}_{\min} = -5$  dB and  $\text{SNR}_{\max} = 5$  dB, in steps of 2.5 dB. For each SNR, 100 statistically independent trials were considered. For each trial,  $K = 100$  samples were generated, the IML algorithm was runned, and the bit error rates (BER) were evaluated by counting the errors; also, the square-error of the channel estimate,  $\|\hat{\mathbf{H}} - \mathbf{H}\|^2$  was computed; here,  $\hat{\mathbf{H}} \equiv \mathbf{P}\hat{\mathbf{Q}}_{ML}$ , recall (4). In figure 1, we plot the mean square error of the channel estimate  $\hat{\mathbf{H}}$  thus obtained: solid line with squares. For comparison, the dashed line with circles denotes the Cramer-Rao Bound (CRB) for

$\mathbf{H}$ , assuming that all the transmitted symbols are known. The curves are quite close. Figures 2 and 3 (solid lines with

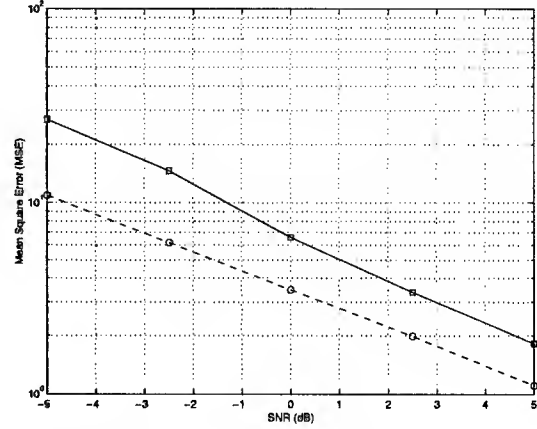


Figure 1: Mean-Square Error: Proposed (solid,square) and Cramer-Rao Bound with  $\theta$  known (dashed,circle)

squares) display the mean BERs obtained for user 1 and user 2, respectively. For comparison, the dashed lines with circles refer to the BERs obtained by linear equalizers based on the true channel matrix  $\mathbf{H}$ , (rows of the pseudo-inverse of  $\mathbf{H}$ ), followed by simple recombination (mean value) of the echos prior to the slicer.

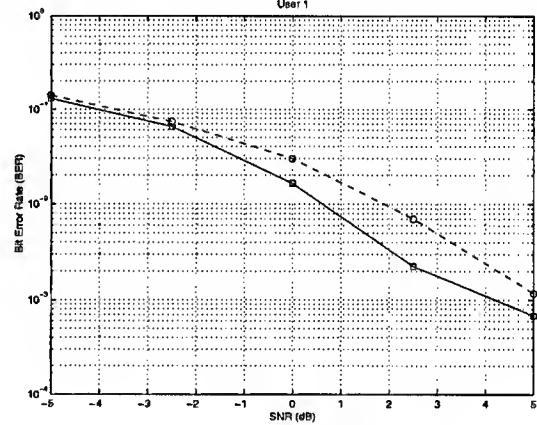


Figure 2: BER 1: Proposed (solid,square), linear equalizer with  $\mathbf{H}$  known (dashed,circle)

#### 5. CONCLUSIONS

We proposed a semi-blind method for separating linear convolutive mixtures of sources when their statistical description is known only up to the 2nd order. We developed an iterative technique (IML algorithm) which permits to compute the joint ML estimator for the orthogonal mixing matrix  $\mathbf{Q}$  and user signals  $\theta$ , in the whitened data space. The IML algorithm respects both the algebraic and temporal constraints on the pair of unknowns  $(\mathbf{Q}, \theta)$ . Future work includes a detailed study on the convergence properties of



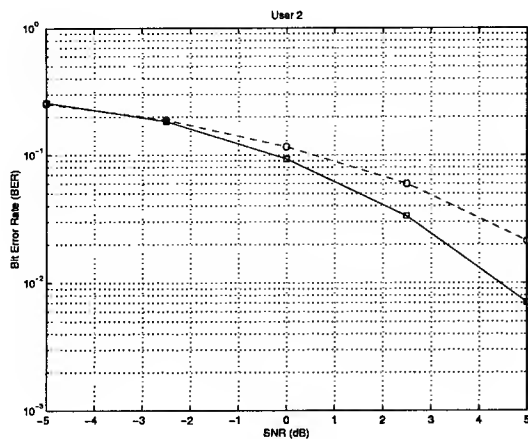


Figure 3: BER 2: Subspace (solid, square), linear equalizer with  $H$  known (dashed, circle)

the IML algorithm (and its optimization subproblems), and evaluation of the computational complexity of both 2 steps (per iteration).

## REFERENCES

- [1] S. Talwar, M. Viberg, and A. Paulraj, "Blind Estimation of Multiple Co-Channel Digital Signals Using an Antenna Array," *IEEE Signal Processing Letters*, vol. 1, no. 2, pp. 29-31, February 1994
- [2] V. Barroso, J. M. F. Moura, and J. Xavier, "Blind Array Channel Division Multiple Access (AChDMA) for Mobile Communications," *IEEE Transactions on Signal Processing*, vol. 46, pp. 737-752, March 1998
- [3] B. Halder, B. Ng, A. Paulraj, and T. Kailath, "Unconditional Maximum Likelihood Approach for Blind Estimation of Digital Signals," in *International Conference on Acoustics, Speech and Signal Processing (ICASSP'96)*, vol. 2, pp. 1081-1084, 1996
- [4] A. Van der Veen, S. Talwar, and A. Paulraj, "A Subspace Approach to Blind Space-Time Signal Processing for Wireless Communication Systems," *IEEE Transactions on Signal Processing*, vol. 45, no. 1, pp. 173-190, January 1997
- [5] J. Xavier and V. Barroso, "Blind source separation, ISI cancellation and carrier phase recovery in SDMA systems for mobile communications," *Wireless Personal Communications*, vol. 10, pp. 35-76, Kluwer Academic Publishers, June 1999
- [6] A. Ghorokov and P. Loubaton, "Subspace based techniques for blind separation of convolutive mixtures with temporally correlated mixtures," *IEEE Transactions on Circuits and Systems - I: Fundamental Theory and Applications*, vol. 44, pp. 813-820, no. 9, September 1997
- [7] R. Bhatia, *Matrix Analysis*, Springer-Verlag New-York, Inc.
- [8] G. H. Golub and C. F. Van Loan, *Matrix Computations*, Baltimore, MD: John Hopkins University Press
- [9] D. Luenberger, *Linear and Nonlinear Programming*, 2nd ed., Addison Wesley

# TECHNIQUES FOR BLIND SOURCE SEPARATION USING HIGHER-ORDER STATISTICS

Ziauddin M. Kamran, A. Rahim Leyman

School of Electrical and Electronic Engineering  
Nanyang Technological University  
Nanyang Avenue, Singapore 639798  
{pm0538251, earleyman}@ntu.edu.sg

Karim Abed-Meraim

ENST/TSI, 46, Rue Barrault  
75634 Paris Cedex 13  
France  
abed@tsi.enst.fr

## ABSTRACT

The blind source separation (BSS) problem consists of the recovery of a set of statistically independent source signals from a set of measurements that are mixtures of the sources when nothing is known about the sources and the mixture structure. This paper considers the separation and estimation of independent sources from their instantaneous linear mixed observed data. The concept of higher-order moment and higher-order time-frequency distribution matrices are also introduced. In practice, separation can be achieved by using suitable second-order statistics (SOS) and/or higher-order statistics (HOS). Computationally feasible implementations are presented based on joint diagonalization of the moment matrices and matrices of the principal slices of the time-multifrequency domain of support of the moment-based Wigner trispectrums. The latter approach allows separation of the sources with nonstationarity properties. Simulation results are given to demonstrate the effectiveness of the proposed approaches.

## 1. INTRODUCTION

In BSS the problem is how to recover independent sources given the sensor outputs in which the sources have been mixed in an unknown channel. A number of applications require the extraction of a set of signals which are not directly accessible. Instead, this extraction must be carried out from another set of measurements which were generated as mixtures of the initial set. Since usually neither the original signals - called *sources* - nor the mixing transformation are known, this is certainly a challenging problem of multichannel blind estimation. This problem is encountered in a wide range of application fields, such as array processing, communications, biomedical signal processing, image processing, and speech processing. While ill-

defined in some situations, BSS becomes a well defined problem in the context of multiple-sensor signal processing. Thus far, numerous approaches have been proposed and implemented to this problem by using HOS based approaches [1], [2], and SOS based approaches [3], [4]. Firstly, we introduce a BSS technique based on a joint diagonalization of several fourth-order moment matrices. In this case, we consider stationary sources. Although most of the approaches are successful under certain assumed conditions, one common limitation involving with them is that they are applicable only for stationary sources. In practical applications, nonstationary processes are frequently encountered in radar, sonar, and communication systems. In contrast to BSS approaches using these techniques, we also propose another approach to take advantage explicitly of the nonstationary property of the signals to be separated. This is accomplished by resorting to the powerful tool of time-frequency ( $t - f$ ) signal representations. This approach for BSS exploits the fourth-order moment spectra based  $t - f$  distributions of the array output. Recently,  $t - f$  distributions have been applied for BSS problems [4]. Simulation examples are presented to demonstrate the effectiveness of the proposed approaches.

## 2. PRELIMINARIES: PROBLEM STATEMENT AND TERMINOLOGY

### 2.1. Problem Description

The goal of BSS can be briefly stated as recovering a set of  $n$  zero-mean statistically independent source signals  $\mathbf{s}(t) = [s_1(t), \dots, s_n(t)]^T$  from a set of  $m$  instantaneous linear mixtures  $\mathbf{x}(t) = [x_1(t), \dots, x_m(t)]^T$ , which are the *observed signals* or *sensor output*. In matrix form, this problem leads to the following data model

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{n}(t) \quad (1)$$

where  $\mathbf{n}(t) = [n_\pi(t), \dots, n_m(t)]^T$  is a  $m \times 1$  additive noise vector whose elements are modeled as stationary, spatially and temporally white, zero-mean complex random processes, and independent of the source signals. That is,  $E[\mathbf{n}(t+\tau)\mathbf{n}^*(t)] = \sigma\delta(\tau)\mathbf{I}$ , where  $\delta(\tau)$  is the Kronecker delta,  $\mathbf{I}$  denotes the identity matrix,  $\sigma$  is the noise power at each sensor, superscript  $*$  denotes conjugate transpose of a vector, and  $E[\cdot]$  is the statistical expectation operator. The unknown  $m \times n$  complex matrix  $\mathbf{A}$  is the full column rank *mixing matrix* or *parameter matrix* that characterizes the medium or channel. The power of the sources is, in principle, arbitrary since a scalar factor can be swapped between any source and its associated column in the mixing matrix without altering the measurements. These well-known facts constitute a basic indeterminacy in BSS [2]. At best, mixing matrix  $\mathbf{A}$  can be identified up to a permutation and scaling of its columns. Therefore, nothing prevents us from further assuming that the sources are unit power signals,  $E[s_i(t)]^2 = 1$  for  $1 \leq i \leq n$ , so that the dynamic range of the sources is accounted for by the magnitude of the corresponding column of  $\mathbf{A}$ . It should be noted, however, that this is merely a convention.

## 2.2. Higher-Order Moments

Let  $x_1(t), \dots, x_m(t)$  be  $m$  random processes and its fourth-order moment sequence can be defined as

$$\begin{aligned} \text{Moment}[x_i(t), x_j^*(t+\tau_1), x_k(t+\tau_2), x_l^*(t+\tau_3)] \\ = \text{MOM}_{x_i x_j^* x_k x_l^*}(\tau_1, \tau_2, \tau_3) \\ = E[x_i(t)x_j^*(t+\tau_1)x_k(t+\tau_2)x_l^*(t+\tau_3)] \end{aligned} \quad (2)$$

where  $1 \leq i, j, k, l \leq m$ . For finite fourth-order moments, we also define a moment set denoted by

$$\mathcal{Q}_x(\tau_1, \tau_2, \tau_3) \triangleq \{\text{MOM}_{x_i x_j^* x_k x_l^*}(\tau_1, \tau_2, \tau_3)\} \quad (3)$$

where  $1 \leq i, j, k, l \leq m$ . We assume that there exists consistent estimate of  $\mathcal{Q}_x(\tau_1, \tau_2, \tau_3)$ . We also assume the source signal vector  $\mathbf{s}(t)$  is a stationary random multivariate process with

$$E[\mathbf{s}(t+\tau)\mathbf{s}^*(t)] = \text{diag}[\rho_{11}(\tau), \dots, \rho_{nn}(\tau)] \quad (4)$$

where  $\text{diag}[\cdot]$  is the diagonal matrix formed with the elements of its vector valued argument, and  $\rho_{ii}(\tau) = E[s_i(t+\tau)s_i^*(t)]$  denotes the autocovariance of  $s_i(t)$ . Since sources are uncorrelated, we define source auto-moments as

$$\nu_p(\tau_1, \tau_2, \tau_3) \triangleq \text{MOM}_{s_p s_p^* s_p s_p^*}(\tau_1, \tau_2, \tau_3) \quad (5)$$

where  $1 \leq p \leq n$ . For notational convenience, we denote  $\nu_p(\tau_1, \tau_2, \tau_3)$  as  $\nu_p$ . We consider that auto-moments of sources  $\nu_p$  exist  $\forall \tau_1, \tau_2, \tau_3$ .

## 2.3. Higher-Order Moment and Spectra Based Time-Frequency Distributions

Time-frequency distributions have proven useful for analyzing a variety of signals and systems. In particular, if the frequency content is time varying as in nonstationary signals, then this approach is quite attractive. An infinite number of  $t-f$  distributions of a signal  $x(t)$ , can be generated from a unified framework using Cohen's general class formulation [5],

$$\begin{aligned} P_x(t, f) = \int_{\tau} \int_{\Omega} \int_u \Phi(\Omega, \tau) x^*(u - \tau/2) x(u + \tau/2) \\ \times \exp(j2\pi u \Omega) \exp(-j2\pi t \Omega) \\ \times \exp(-j2\pi f \tau) d\tau d\Omega du \end{aligned} \quad (6)$$

where  $t$  and  $f$  represent the time index and the frequency index, respectively. The kernel  $\Phi(\Omega, \tau)$  characterizes the distribution and is a function of both time and lag variables. For the usual case of signal independent kernel  $\Phi(\Omega, \tau)$  all members of Cohen's general class are bilinear with respect to signal. This bilinearity is necessary to obtain a second-order spectral analysis of the signal. However, bilinear representation can not give information about the temporal evolution of the higher- (than second) order spectrum of the signal. New representations that could illustrate the time-varying higher-order spectral information of the signal under analysis need to be defined. They should incorporate a higher-order nonlinearity in their formulation. The unified approach given by Cohen to most  $t-f$  representations can be generalized to the case of higher-order moment spectra. A general class of higher-order moment and spectra based  $t-f$  distributions is proposed in [6] as extension of Cohen's general class of distributions to the higher-order spectral domains. Let  $x_1(t), \dots, x_m(t)$  be  $m$  random processes. Define moment-based fourth-order Wigner distribution or Wigner trispectrum (WT) by

$$\begin{aligned} W_{x_i x_j^* x_k x_l^*}(t, \mathbf{f}) \\ = \int_{\tau} \text{MOM}_{x_i x_j^* x_k x_l^*}(t, \tau) \exp(-j2\pi \mathbf{f}^T \tau) d\tau \end{aligned} \quad (7)$$

where  $1 \leq i, j, k, l \leq m$ ;  $t$  and  $\mathbf{f} = [f_1, f_2, f_3]^T$  represent time index and multifrequency index respectively;  $\tau = [\tau_1, \tau_2, \tau_3]^T$  and  $d\tau = d\tau_1 d\tau_2 d\tau_3$ . Here, we define  $\text{MOM}_{x_i x_j^* x_k x_l^*}(t, \tau)$  as follows:

$$\begin{aligned} \text{MOM}_{x_i x_j^* x_k x_l^*}(t, \tau) \\ = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T x_i(t+a) x_j^*(t+b) x_k(t+c) x_l^*(t+d) \end{aligned} \quad (8)$$

where  $a, b, c$ , and  $d$  are obtained from the following requirements:

- (1)  $b - a = \tau_1$ ,  $c - a = \tau_2$ , and  $d - a = \tau_3$  and
- (2) the "lag centering" condition,  $a + b + c + d = 0$ .

These two requirements lead to  $a = -(\tau_1 + \tau_2 + \tau_3)/4$ ,  $b = (3\tau_1 - \tau_2 - \tau_3)/4$ ,  $c = (3\tau_2 - \tau_1 - \tau_3)/4$ , and  $d = (3\tau_3 - \tau_1 - \tau_2)/4$ . Then the moment-based fourth-order time-frequency distribution is formulated as

$$\begin{aligned} P_{x_i x_j^* x_k x_l^*}(t, \mathbf{f}) \\ = \int_{\tau} \int_{\Omega} \int_u \Phi(\Omega, \tau) Mom_{x_i x_j^* x_k x_l^*}(u, \tau) \exp(j2\pi u \Omega) \\ \times \exp(-j2\pi t \Omega) \exp(j2\pi \mathbf{f}^T \tau) d\tau d\Omega du \end{aligned} \quad (9)$$

where the three dimensional kernel  $\Phi(\Omega, \tau)$  characterizes the distribution. The numerical implementation of the WT requires computation of the three-dimensional FFT. Computation of the WT at least at the signal rate for high temporal resolution and for a frequency resolution of  $\frac{1}{N}$  demands  $\mathcal{O}(N^4 \log_2 N)$  operations in WT, where  $N$  is the length of data window. As proposed by Fonollosa *et al.* [7], a computationally feasible implementation of the WT can be obtained by considering two-dimensional slices of the time multi-frequency domain of support of the WT. Each slice corresponds to the plane defined by the temporal axis and one frequency line that represents jointly all three frequency axes of the four dimensional WT. Care must be taken to choose this frequency line appropriately to contain the information of interest for every particular application. Among all possible slices, we choose the principal slice corresponding to the temporal axis and principal diagonal of the higher-order spectra. The principal slice is thus the plane corresponding to  $f_\pi = f_2 = -f_3$  in the time-multifrequency domain of the WT. Our approach to blind identification exploits the principal sliced versions of the WTs defined as the sliced Wigner trispectrums (SWTs) of the array output. From (7) we define the SWT as

$$\begin{aligned} SW_{x_i x_j^* x_k x_l^*}(t, f) &= W_{x_i x_j^* x_k x_l^*}(t, \mathbf{f}) \Big|_{f_1=f_2=-f_3=f} \\ &= \int_{\tau_1} \int_{\tau_2} \int_{\tau_3} Mom_{x_i x_j^* x_k x_l^*}(t, \tau) \\ &\quad \times \exp(-j2\pi f(\tau_1 + \tau_2 - \tau_3)) d\tau_1 d\tau_2 d\tau_3. \end{aligned} \quad (10)$$

The bilinear dependence on the signal of Cohen's class is substituted by a multilinear form in the WT. Consequently, cross-terms are more numerous in WT than they are in the bilinear representations, which, if allowed to pass into WT, can reduce auto-component

resolution and obscure the true signal features. To reduce the cross-terms of the principal slice of the WT effectively, we apply Choi-Williams (CW) exponential kernel to the SWT and hence, we define the sliced reduced interference trispectrum distribution (SRID) as

$$\begin{aligned} SRID_{x_i x_j^* x_k x_l^*}(t, f) \\ = \int_{\Omega} \int_{\tau} \exp(\tau^2 \Omega^2 / \varrho) \left[ \int_{t'} \int_{f'} SW_{x_i x_j^* x_k x_l^*}(t', f') \right. \\ \times \exp(-j2\pi t' \Omega) \exp(-j2\pi f' \tau) dt' df' \Big] \\ \times \exp(j2\pi t \Omega) \exp(j2\pi f \tau) d\Omega d\tau \end{aligned} \quad (11)$$

where  $\varrho$  is the kernel width. We define a SRID set denoted  $\mathcal{B}_x(t, f)$  as

$$\mathcal{B}_x(t, f) \triangleq \{SRID_{x_i x_j^* x_k x_l^*}(t, f)\} \quad (12)$$

where  $1 \leq i, j, k, l \leq m$ . We assume that there exists consistent estimate of  $\mathcal{B}_x(t, f)$ . In this case, we also assume that the source signal vector  $\mathbf{s}(t)$  is a nonstationary random multivariate process with

$$\lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1, T} \mathbf{s}(t + \tau) \mathbf{s}^*(t) = \text{diag}[\gamma_{11}(\tau), \dots, \gamma_{nn}(\tau)] \quad (13)$$

where the autocovariance of  $s_i(t)$  is denoted by  $\gamma_{ii}(\tau) = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1, T} s_i(t + \tau) s_i^*(t)$ . Since sources are independent, the source cross-cumulants or cross-terms of the source SRID vanish. Hence, the source auto-terms are denoted as

$$\kappa_p(t, f) \triangleq SRID_{s_p s_p^* s_p s_p^*}(t, f) \quad (14)$$

where  $1 \leq p \leq n$ . For notational convenience, we denote  $\kappa_p(t, f)$  as  $\kappa_p$ . We also assume that SRIDs of sources  $\kappa_p$  exist  $\forall t, f$ .

### 3. HIGHER-ORDER STATISTICS BASED BSS APPROACHES

Let  $\mathbf{W}$  denotes a  $n \times m$  whitening matrix such that  $\mathbf{W}\mathbf{A} = \mathbf{U}$ , where  $\mathbf{U}$  is a  $n \times n$  unitary matrix. The whitened process still obeys a linear model

$$\mathbf{z}(t) \triangleq \mathbf{W}\mathbf{x}(t) = \mathbf{W}[\mathbf{A}\mathbf{s}(t) + \mathbf{n}(t)] = \mathbf{U}\mathbf{s}(t) + \mathbf{W}\mathbf{n}(t). \quad (15)$$

We form moment matrices  $\mathbf{M}_z^P(\tau_1, \tau_2, \tau_3)$  by the inner product of the moments of the whitened data with an arbitrary  $n \times n$  matrix  $\mathbf{P}$ , i.e.,

$$[\mathbf{M}_z^P(\tau_1, \tau_2, \tau_3)]_{ij} \triangleq \sum_{k=1}^n \sum_{l=1}^n MOM_{z_i z_j^* z_k z_l^*}(\tau_1, \tau_2, \tau_3) P_{lk} \quad (16)$$

where  $1 \leq i, j \leq n$ , and the  $(l, k)$ th component of the matrix  $\mathbf{P}$  is written as  $P_{lk}$ . Since sources are uncorrelated and the fourth-order moments of different elements of the independent and identically distributed (i.i.d.) noise vector,  $MOM_{n_i n_j n_k n_l}(\tau_1, \tau_2, \tau_3) = 0$  [8], where  $1 \leq i, j, k, l \leq m$ , and  $\tau_1 \neq \tau_2 \neq \tau_3$ , the relation (15) yields

$$MOM_{z_i z_j^* z_k z_l^*}(\tau_1, \tau_2, \tau_3) = \sum_{p=1}^n \nu_p u_{ip} u_{jp}^* u_{kp} u_{lp}^* \quad (17)$$

where  $u_{ip}$  denotes the  $(i, p)$ th entry of the matrix  $\mathbf{U}$ . Therefore, the whitened moment matrix can be rewritten as

$$\begin{aligned} \mathbf{M}_z^{\mathbf{P}}(\tau_1, \tau_2, \tau_3) &= \sum_{p=1}^n \nu_p \mathbf{u}_p^* \mathbf{P} \mathbf{u}_p \mathbf{u}_p \mathbf{u}_p^* \quad \forall \mathbf{P} \\ \Rightarrow \mathbf{M}_z^{\mathbf{P}}(\tau_1, \tau_2, \tau_3) &= \mathbf{U} \mathbf{\Lambda}_p^{\mathbf{P}} \mathbf{U}^H \end{aligned} \quad (18)$$

where  $\mathbf{\Lambda}_p^{\mathbf{P}} = \text{diag}[\nu_1 \mathbf{u}_1^* \mathbf{P} \mathbf{u}_1, \dots, \nu_n \mathbf{u}_n^* \mathbf{P} \mathbf{u}_n]$  is a diagonal matrix whose diagonal elements depend on the particular matrix  $\mathbf{P}$  as well as the whitened data vector  $\mathbf{z}(t)$ , superscript  $H$  denotes the complex conjugate transpose of a matrix, and  $\mathbf{u}_p$  denotes  $p$ th column of the matrix  $\mathbf{U}$ .

Similarly we can form the whitened SRID matrices as follows

$$\begin{aligned} \mathbf{S}_z^{\mathbf{P}}(t, f) &= \sum_{p=1}^n \kappa_p \mathbf{u}_p^* \mathbf{P} \mathbf{u}_p \mathbf{u}_p \mathbf{u}_p^* \quad \forall \mathbf{P} \\ \Rightarrow \mathbf{S}_z^{\mathbf{P}}(t, f) &= \mathbf{U} \mathbf{\Lambda}_p^{\mathbf{P}} \mathbf{U}^H \end{aligned} \quad (19)$$

where  $\mathbf{\Lambda}_p^{\mathbf{P}} = \text{diag}[\kappa_1 \mathbf{u}_1^* \mathbf{P} \mathbf{u}_1, \dots, \kappa_n \mathbf{u}_n^* \mathbf{P} \mathbf{u}_n]$  is a diagonal matrix whose diagonal elements depend on the particular matrix  $\mathbf{P}$  as well as the whitened data vector  $\mathbf{z}(t)$ . From (18) and (19) stem the basic idea for eigen-based blind identification using moment matrices and SRID matrices of the observed data, respectively. Thus matrix  $\mathbf{U}$  diagonalizes  $\mathbf{M}_z^{\mathbf{P}}(\tau_1, \tau_2, \tau_3)$  or  $\mathbf{S}_z^{\mathbf{P}}(t, f)$  and the columns of the unknown  $\mathbf{U}$  can be identified to the eigen-vectors of  $\mathbf{M}_z^{\mathbf{P}}(\tau_1, \tau_2, \tau_3)$  or  $\mathbf{S}_z^{\mathbf{P}}(t, f)$  for any matrix  $\mathbf{P}$ . A similar diagonalization of fourth-order cumulant matrices is explored in [1].

We form the sample fourth-order moments  $Q_z(\mathcal{L}_q)$  of the whitened process  $\mathbf{z}(t) = \mathbf{W}\mathbf{x}(t)$  for a fixed set of time lags  $\mathcal{L}_q = \{\tau_q, \tau_{q+1}, \tau_{q+2} \mid 1 \leq q \leq K \text{ and } \tau_q \neq \tau_{q+1} \neq \tau_{q+2}\}$ . To reduce the possibility of having degenerate eigenvalues as well as to reduce the effect of the noise, we compute the  $n$  most significant eigenpairs  $\{\lambda_r', \mathbf{P}_r' \mid 1 \leq r \leq n\}$  from the eigen-structure of  $Q_z(\mathcal{L}_q)$  derived from (18) corresponding to each time lag combination  $\{\tau_q, \tau_{q+1}, \tau_{q+2}\}$  in  $\mathcal{L}_q$ . Similarly, we form the sample SRIDs  $\mathcal{B}_z(t_q, f_q)$  of the whitened process for a fixed set of  $(t_q, f_q)$  points,  $1 \leq q \leq K$ . The

$n$  most significant eigenpairs  $\{\lambda_r'', \mathbf{P}_r'' \mid 1 \leq r \leq n\}$  are computed from the eigen-structure of  $\mathcal{B}_z(t_q, f_q)$  derived from (19) corresponding to each  $(t, f)$  point of a fixed set of  $(t_q, f_q)$  points which correspond to the signal autoterms. The whitening matrix  $\mathbf{W}$  can be obtained from SOS [1], [3], [4]. A unitary matrix  $\mathbf{U}$  is then obtained as joint diagonalizer of the  $K \times n$  matrices obtained from the set  $\mathcal{M}' \triangleq \{\lambda_r' \mathbf{P}_r' \mid 1 \leq r \leq n\}$  corresponding to each  $\{\tau_q, \tau_{q+1}, \tau_{q+2}\}$  in  $\mathcal{L}_q$ , or from the set  $\mathcal{M}'' \triangleq \{\lambda_r'' \mathbf{P}_r'' \mid 1 \leq r \leq n\}$  corresponding to each  $(t, f)$  point of a fixed set of  $(t_q, f_q)$  points which correspond to the signal autoterms. An efficient joint approximate diagonalization technique exists in [1], [3], [9]. The source signals are estimated as  $\hat{\mathbf{s}}(t) = \hat{\mathbf{U}}^H \hat{\mathbf{W}} \mathbf{x}(t)$ , and/or the mixing matrix is estimated as  $\hat{\mathbf{A}} = \hat{\mathbf{W}}^{\#} \mathbf{U}$ , where the superscript  $\#$  denotes the Moore-Penrose pseudoinverse.

#### 4. SIMULATION RESULTS

Computer simulations are conducted to illustrate the performance of the proposed approaches. In the simulations environment, a three-element uniform linear array with half wavelength sensor spacing receives two signals in the presence of white Gaussian noise. The first chirp signal contains a quadratic phase coupling, whereas the second signal is logarithmic chirp signal. The sources arrive from different directions  $\phi_1 = 10^\circ$  and  $\phi_2 = 50^\circ$ . While choosing  $t_q - f_q$  points of the sets of SRIDs  $\mathcal{B}_z(t_q, f_q)$ , we take  $K$  time-frequency high signal to noise ratio (SNR) auto-term points. The performance is evaluated by using mean rejection level  $= \sum_{b \neq a} E[|\hat{\mathbf{A}}^{\#} \mathbf{A}|_{ab}|^2]$  [4], where  $E[|\hat{\mathbf{A}}^{\#} \mathbf{A}|_{ab}|^2]$  measures the ratio of the power of the interference of the  $b$ th source to the power of the  $a$ th estimated source signal. Fig. 1 shows the mean rejection levels of the proposed approaches over the range from -10 dB to 20 dB of SNR. In this case, 10 matrices are considered for joint diagonalization. In Fig. 2, the mean rejection levels are plotted as the function of the number of the jointly diagonalized matrices for SNR = 0 dB. The mean rejection levels are evaluated over 100 Monte-Carlo trials with 512 snapshots. It is apparent that the proposed approaches have clearly succeeded to estimate the unknown mixing matrix. The approach using SRID matrices shows better performance than that of using moment matrices. The excellence of using  $t - f$  representations for the BSS problem is proven again.

#### 5. CONCLUSION

In this paper, we propose BSS approaches using fourth-order moments and its time-frequency representations.

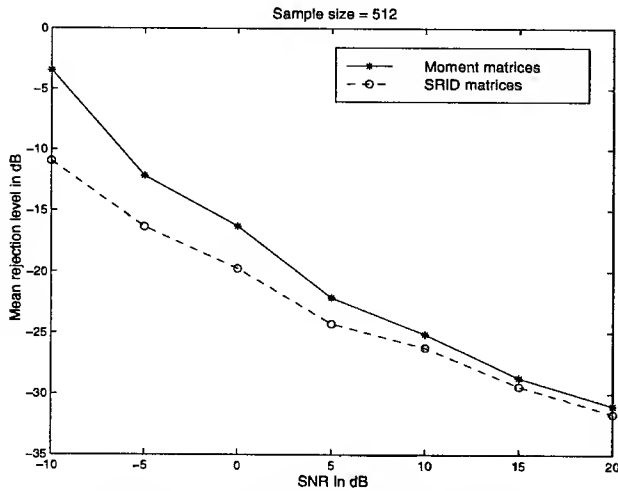


Figure 1. Mean rejection level versus SNR.

These are based on the joint diagonalization of the moment matrices and SRID matrices of the principal slices of time-multifrequency domain of support of the WTs, which allows the whole moment and trispectrum set to be processed with a computational efficiency similar to eigen-based techniques. Moreover, computation of two-dimensional slices of trispectrums demands the same computational complexity and time-frequency resolution as bilinear time-frequency representations. The important distortion introduced by cross-terms of the principal slice of the WT is reduced effectively by the application of the exponential kernel of the CW distribution. Hence overwhelming complexity due to a product in the fourth-order ambiguity domain with the fourth-dimensional extension of the CW kernel can be avoided. Numerical experiments are presented to assess the effectiveness of the proposed approaches. The approach using SRID matrices shows better performance than that of using moment matrices. This is because of the fact that the effect of spreading the noise power while localizing the source energy in the  $t-f$  domain amounts to increasing the robustness of the choosing SRID matrices with respect to noise.

## 6. REFERENCES

- [1] J.-F. Cardoso and A. Souloumiac, "Blind beamforming for non Gaussian signals," *IEE Proceedings-F*, vol. 140, no. 6, pp. 362-370, December 1993.
- [2] L. Tong, R. Liu, V. Soon, and Y. Huang, "Indeterminacy and identifiability of blind identification," *IEEE Trans. on Circuits and Systems*, vol. 38, no. 5, pp. 499-509, May 1991.

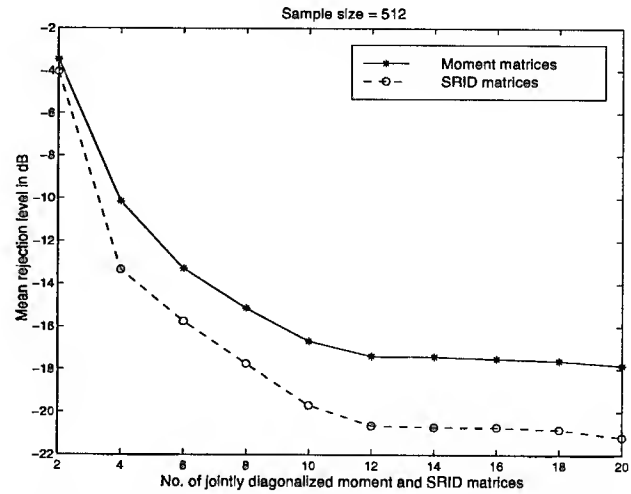


Figure 2. Mean rejection level versus number of jointly diagonalized moment and SRID matrices.

- [3] A. Belouchrani, K. Abed-Meraim, J.-F. Cardoso, and E. Moulines, "A blind source separation technique using second order statistics," *IEEE Trans. on Signal Processing*, vol. 45, no. 2, pp. 434-444, February 1997.
- [4] A. Belouchrani and M. G. Amin, "Blind source separation based on time-frequency signal representations," *IEEE Trans. on Signal Processing*, vol. 46, no. 11, pp. 2888-2897, November 1998.
- [5] L. Cohen, "Time-frequency distributions - A review," *Proceedings of the IEEE*, vol. 77, no. 7, pp. 941-981, July 1989.
- [6] J. R. Fonollosa and C. L. Nikias, "Wigner higher order moment spectra: Definition, properties, computation and application to transient signal analysis," *IEEE Trans. on Signal Processing*, vol. 41, no. 1, pp. 245-266, January 1993.
- [7] J. R. Fonollosa and C. L. Nikias, "Analysis of finite-energy signal using higher-order moments and spectra-based time-frequency distributions," *Signal Processing*, vol. 36, no. 3, pp. 315-328, April 1994.
- [8] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, New York: McGraw-Hill, 1991.
- [9] A. Bunse-Gerstner, R. Byers, and V. Mehrmann, "Numerical methods for simultaneous diagonalization," *SIAM Journal of Matrix Analysis and Applications*, vol. 14, pp. 927-949, 1993.

# JOINT-DIAGONALIZATION OF CUMULANT TENSORS AND SOURCE SEPARATION

Eric MOREAU

MS-GEISSY, ISITV, av. G. Pompidou, BP 56,  
F-83162 La Valette du Var Cedex, France  
e-mail: moreau@isitiv.univ-tln.fr

## ABSTRACT

In this paper, we extend the results leading to the popular JADE and STOTD algorithms to cumulants of any order greater than or equal to three. We first exhibit a new contrast function which constitutes an unified framework for the underlying contrasts of JADE and STOTD which thus appear as particular cases. Then we generalize the link between these new contrasts and a joint-diagonalization criteria. Moreover for the generalized JADE's contrast, the analytical optimal solution in the case of two sources is derived and shown to keep the same simple expression whatever the cumulant order. Finally, some computer simulations illustrate the potential advantage one can take of considering statistics of different orders for the joint-diagonalization of cumulant matrices.

## 1. INTRODUCTION

The problem of source separation has found numerous solutions in the past decade. Beginning with the originate works of Héroult and Jutten, see [1] and references therein, who have proposed an adaptive (on-line) algorithm, three of the most important contributions are provided by Comon [2], Cardoso and Souloumiac [3] and De Lathauwer [5]. These later solutions are block (off-line) algorithms which are both closely related to contrast functions (also simply called contrasts). Such contrasts were introduced and defined in [2] and have recently found a generalization in [4]. The algorithm presented in [2] is called ICA for "Independent Component Analysis". The algorithm presented in [3] is called JADE for "Joint Approximate Diagonalization of Eigen-matrices" and the one presented in [5] is called STOTD for "Simultaneous Third Order Tensor Diagonalization".

The JADE's and STOTD's underlying contrasts only take into consideration fourth order cumulants on the contrary to the ICA's one which remains available whatever the order of cumulants is since it is greater

than or equal to three. On the other hand, the fourth order JADE's and STOTD's contrasts have also found interesting interpretations in terms of a joint-diagonalization criteria. For the JADE's contrast (resp. the STOTD' contrast) it is a joint diagonalization criterion (maximized w.r.t. a unitary matrix) of some cumulant matrices sets (resp. of some cumulant third order tensors sets). These links are the keys for the derivation of the practical JADE and STOTD algorithms.

In this paper, we are mainly interested in generalizing the underlying contrasts of JADE and STOTD and their links with joint-diagonalization criteria involving cumulants of any order greater than or equal to three. The main interests are then to be able to choose the cumulants order or to combine statistical information of different orders.

## 2. PROBLEM FORMULATION

Signals emitted from different sources are observed thanks to

$$\mathbf{x}(n) = \mathbf{G}\mathbf{a}(n) \quad (1)$$

where  $n \in \mathbb{Z}$  is the discrete time,  $\mathbf{a}(n)$  the  $(N, 1)$  vector of  $N \neq 2$  unobservable *real* input signals  $a_i(n)$ ,  $i \in \{1, \dots, N\}$ , called sources,  $\mathbf{x}(n)$  the  $(N, 1)$  vector of observed signals  $x_i(n)$ ,  $i \in \{1, \dots, N\}$  and  $\mathbf{G}$  the  $(N, N)$  square mixing matrix assumed *invertible*. For clarity, in this article we restrict our attention to the case of real signals and mixtures although the following derivations might be easily extended to complex ones.

Further, the following assumptions are considered

A1. "*Independence*" The sources  $a_i(n)$ ,  $i \in \{1, \dots, N\}$ , are zero-mean, unit power and statistically mutually independent;

A2. "*Stationarity*"  $a_i(n)$ ,  $i \in \{1, \dots, N\}$ , are random signals stationary up to order under consideration, i.e.  $\forall i \in \{1, \dots, N\}$ , the cumulant  $\text{Cum} \underbrace{[a_i(n), \dots, a_i(n)]}_{R \times}$

is an independent function of  $n$ , denoted by  $C_R[a_i]$ ;

moreover at most one of the cumulants  $C_R[a_i]$ ,  $i \in \{1, \dots, N\}$ , is null.

It is important now to introduce the notion of white vectors [2][3] because of its use as a first transformation in the JADE and STOTD algorithms. A vector  $\mathbf{z}(n)$  of random signals is said to be (spatially) white if its covariance matrix  $\mathbf{R}_z = E[\mathbf{z}\mathbf{z}^T]$  equals the identity. The first (second order) transformation is then defined as a whitening of the observation vector  $\mathbf{x}(n)$ . This is done by applying a whitening matrix  $\mathbf{B}$  in such a way that  $\mathbf{B}\mathbf{G} = \mathbf{V}$  where  $\mathbf{V}$  is a unitary matrix, i.e.  $\mathbf{V}\mathbf{V}^T = \mathbf{I}$ . Hence, after the whitening transformation, the new "observed" vector reads

$$\mathbf{x}_b(n) = \mathbf{B}\mathbf{x}(n) = \mathbf{V}\mathbf{a}(n). \quad (2)$$

The blind source separation problem consists now in estimating a unitary matrix  $\mathbf{H}$  in such a way that the vector

$$\mathbf{y}(n) = \mathbf{H}\mathbf{x}_b(n) \quad (3)$$

restores one of the different (possibly noisy) sources on each of its different components.

Because the sources are inobservable and the mixture is unknown, the exact power and order of each sources can not be recovered. It is the reason why the separation is said to be achieved when the global unitary matrix  $\mathbf{S}$  defined as

$$\mathbf{S} = \mathbf{H}\mathbf{V} \quad (4)$$

reads

$$\mathbf{S} = \mathbf{D}\mathbf{P} \quad (5)$$

where  $\mathbf{D}$  is an invertible diagonal matrix (here with unit modulus components) corresponding to arbitrary attenuations (here signs) for the restored sources and  $\mathbf{P}$  a permutation matrix corresponding to an arbitrary order of restitution. According to (3), (1) and (4) the output vector can be written as

$$\mathbf{y}(n) = \mathbf{S}\mathbf{a}(n). \quad (6)$$

Because of the stationarity assumption, the explicit dependence of sources, observations and outputs vectors with the discrete time  $n$  will be now omitted whenever no confusion is possible.

Let us define some notations which will be useful in the following. Let  $\mathcal{A}$  be the set of random vectors satisfying assumptions A1 and A2. Let  $\mathcal{U}$  be the set of unitary matrices. The subset of  $\mathcal{U}$  of matrices  $\mathbf{S}$  of the form (5) is denoted by  $\mathcal{P}$  and the subset of  $\mathcal{P}$  of diagonal matrices is denoted by  $\mathcal{D}$ . Finally the set of random vector  $\mathbf{y}(n)$  built from (6) where  $\mathbf{a}(n) \in \mathcal{A}$  and  $\mathbf{S} \in \mathcal{U}$  is denoted by  $\mathcal{Y}_u$ .

### 3. A NEW CONTRAST

A contrast is usually a function of the output of the separating system. As defined in [2][4], its (global) maximization arguments yield a separating solution, i.e. a matrix  $\mathbf{H}$  such that the global matrix  $\mathbf{S}$  can be factored as in (5). The definition in [4] is now recalled for readers convenience.

**Definition 1** A contrast on  $\mathcal{Y}_u$  is a multivariate mapping  $\mathcal{I}(\cdot)$  from  $\mathcal{Y}_u$  to the real set which satisfies the following three requirements:

- R1.  $\forall \mathbf{y} \in \mathcal{Y}_u, \forall \mathbf{D} \in \mathcal{D}, \mathcal{I}(\mathbf{D}\mathbf{y}) = \mathcal{I}(\mathbf{y})$ ;
- R2.  $\forall \mathbf{a} \in \mathcal{A}, \forall \mathbf{S} \in \mathcal{U}, \mathcal{I}(\mathbf{S}\mathbf{a}) \leq \mathcal{I}(\mathbf{a})$ ;
- R3.  $\forall \mathbf{a} \in \mathcal{A}, \forall \mathbf{S} \in \mathcal{U}, \mathcal{I}(\mathbf{S}\mathbf{a}) = \mathcal{I}(\mathbf{a}) \Rightarrow \mathbf{S} \in \mathcal{P}$ .

Historically, one of the first contrast can be found in [2]. Other examples of contrasts can be found in [4]. But of primary importance with our purpose, in [3] and [5] two contrasts involving both cross-cumulants and auto-cumulants have been proposed. The JADE's contrast [3] reads

$$\mathcal{J}(\mathbf{y}) = \sum_{i_1, i_2, i_3=1}^N (\text{Cum}[y_{i_1}, y_{i_1}, y_{i_2}, y_{i_3}])^2 \quad (7)$$

and the STOTD's contrast [5] reads

$$\mathcal{S}(\mathbf{y}) = \sum_{i_1, i_2=1}^N (\text{Cum}[y_{i_1}, y_{i_1}, y_{i_1}, y_{i_2}])^2. \quad (8)$$

We now generalize these functions to cumulants of any order greater than or equal to three, including them in a common generalized family of contrasts. This is given by the following proposition.

**Proposition 1** Let  $R$ ,  $R_1$  and  $R_2$  three integers such that  $R = R_1 + R_2$ ,  $R \geq 3$  and  $2 \leq R_1 \leq R$ , using the notation

$$C_R^{R_1}[\mathbf{y}] = \text{Cum}[\underbrace{y_{i_1}, \dots, y_{i_1}}_{R_1 \times}, \underbrace{y_{i_2}, \dots, y_{i_{R_2+1}}}_{R_2 \text{ terms}}] \quad (9)$$

the function

$$\mathcal{G}_R^{R_1}(\mathbf{y}) = \sum_{i_1, i_2, \dots, i_{R_2+1}=1}^N |C_R^{R_1}[\mathbf{y}]|^2 \quad (10)$$

is a contrast on  $\mathcal{Y}_u$ , i.e. for white vectors  $\mathbf{y}$ .

By definition if  $R_2 = 0$  no corresponding additional terms are considered in the cumulant in (9) which then corresponds to an auto-cumulant. Hence in expression (10), if  $R_2 = 0$  no sum over  $i_2$  is considered. Let us



remark that if  $R_2 = 0$ , we get the ICA's contrast. On the other hand, considering  $R_1 = R_2 = 2$  we have the JADE's contrast while considering  $R_1 = 3$  and  $R_2 = 1$  we have the STOTD's contrast. All other values of  $R_1$  and  $R_2$  yield a new contrast.

#### 4. LINKS WITH JOINT-DIAGONALIZATION

We now show a link between the above contrast  $\mathcal{G}_R^{R_1}(\mathbf{y})$  and a joint-diagonalization criterion of  $R_1$ -order symmetric tensors by an unitary matrix. Such a joint-diagonalization criterion is defined according to

**Definition 2** *Considering a set of  $T$  square and symmetric  $R_1$ -order tensors  $\mathbf{T}(m)$ ,  $m = 1, \dots, T$  denoted by  $\mathcal{T}$ . A joint-diagonalizer of this set is an unitary matrix that maximizes the function*

$$\mathcal{D}(\mathbf{H}, \mathcal{T}) = \sum_{m=1}^T \left( \sum_i |T_{i, \dots, i}^H(m)|^2 \right) \quad (11)$$

where

$$T_{i, \dots, i}^H(m) = \sum_{n_1, \dots, n_{R_1}} H_{i, n_1} \cdots H_{i, n_{R_1}} T_{n_1, \dots, n_{R_1}}(m). \quad (12)$$

Now the equivalence between the contrast  $\mathcal{G}_R^{R_1}(\mathbf{y})$  and a joint-diagonalization criterion of  $R_1$ -order tensors can be stated according to the following proposition:

**Proposition 2** *With  $R \geq 3$  and  $2 \leq R_1 \leq R$ , let  $\mathcal{C}_R^{R_1}$  be the set of  $T = N^{R-R_1}$  tensors of order  $R_1$*

$$\mathbf{T}(i_{R_1+1}, \dots, i_R) = (T_{i_1, \dots, i_{R_1}}(i_{R_1+1}, \dots, i_R))$$

defined as

$$T_{i_1, \dots, i_{R_1}}(i_{R_1+1}, \dots, i_R) = \text{Cum}[x_{i_1}, \dots, x_{i_R}]. \quad (13)$$

Then, if  $\mathbf{H}$  is a unitary matrix, we have

$$\mathcal{D}(\mathbf{H}, \mathcal{C}_R^{R_1}) = \mathcal{G}_R^{R_1}(\mathbf{H}\mathbf{x}). \quad (14)$$

For  $R_1 = 2$  (resp.  $R_1 = 3$ ), this proposition 2 is a generalization of one result in [3] (resp. in [5]) to cumulants of any order greater than or equal to three. For all other values of  $R_1$ , this corresponds to new results.

According to the above proposition, we can now choose the order of cumulants (greater than or equal to three) for the joint-diagonalization of matrices or tensors. In particular third order cumulants can be used leading to the joint-diagonalization of  $N$  matrices or to the diagonalization of only one tensor of order three.

However even if it is sufficient to joint-diagonalize tensors of cumulants of a given order, one can find interest in combining cumulants of different orders. For example this can lead to algorithms that are more robust w.r.t. the statistics of sources. For example one can combine third and fourth order cumulants. If third order (resp. fourth order) cumulants of the unknown sources vanish then the other fourth order (resp. third order) ones can be directly used. In the unfortunate case where both third and fourth order cumulants of the sources vanish, then one has to consider cumulants of greater order. Moreover such combination can be useful for an independent component analysis goal when one is not sure that the available data conform the initial model. Indeed in such a case cross cumulants of all orders have to be canceled.

Now given the order of tensors to be joint-diagonalized, we show that cumulants of different orders can be considered altogether. This is given according to the following proposition:

**Proposition 3** *Let  $\gamma_1, \dots, \gamma_m$  be  $m \in \mathbb{N}^*$  real non negative constants with at least one non zero. Let  $S_1, \dots, S_m$  be  $m$  integers such that  $3 \leq S_1 < \dots < S_m$ . Finally, with  $R_1 \leq S_1$ , let*

$$\sqrt{\gamma_i} \mathcal{C}_{S_i}^{R_1} = \{\sqrt{\gamma_i} \mathbf{T}(i_{R_1+1}, \dots, i_{S_i})\}$$

be  $m$  sets of  $R_1$  order tensors  $\mathbf{T}(\cdot)$  of  $S_i$  order cumulants as defined in proposition 2. Then, if  $\mathbf{H}$  is a unitary matrix, we have

$$\mathcal{D}(\mathbf{H}, \bigcup_{i=1}^m \sqrt{\gamma_i} \mathcal{C}_{S_i}^{R_1}) = \sum_{i=1}^m \gamma_i \mathcal{G}_{S_i}^{R_1}(\mathbf{H}\mathbf{x}). \quad (15)$$

Now since it is well-known that a (non zero) non negative linear combination of contrasts is also a contrast then the joint-diagonalization of tensors of cumulants of mixed orders is again a sufficient condition for separation.

#### 5. GENERALIZATION OF THE JADE ALGORITHM

The JADE algorithm is based on Jacobi optimization. This means that the maximization of the criterion under consideration is realized through a sequence of plane (or Givens) rotations as initiated in [2]. Each plane rotation works on a pair of the output vector  $\mathbf{y}(n)$  and one "sweep" or iteration consists in processing the outputs through all the  $N(N-1)/2$  possible pairs. Hence the  $N$ -dimensional problem is reduced to  $N(N-1)/2$  problems of dimension 2. One of the main advantages is that the 2-dimensional problem is simpler and often admits an analytical solution. Thus let

us now consider the only 2-dimensional problem where a plane rotation has to be determined. In the following, we parameterize it as

$$\mathbf{H} = \begin{pmatrix} \cos \theta & \sin \theta \\ -\sin \theta & \cos \theta \end{pmatrix}. \quad (16)$$

We only consider the joint-diagonalization of matrices and then the maximization of  $\mathcal{G}_R^2(\mathbf{y})$ . For  $N = 2$ , it is easily seen that  $\mathcal{G}_R^2(\mathbf{y})$  can be written as

$$\mathcal{G}_R^2(\mathbf{y}) = \mathbf{u}_\theta^T \mathbf{A}_R \mathbf{u}_\theta \quad (17)$$

where  $\mathbf{u}_\theta^T = (\cos 2\theta \quad \sin 2\theta)$  and where  $\mathbf{A}_R = (A_{R,i,j})$  is a  $(2, 2)$  real symmetric matrix defined according to

$$\begin{aligned} A_{R,1,1} &= t_{R,1}; \\ A_{R,1,2} &= t_{R,4}; \\ A_{R,2,2} &= \frac{1}{2}t_{R,1} + 2t_{R,2} + t_{R,3}. \end{aligned} \quad (18)$$

with, using (13) with  $R_1 = 2$ ,

$$\begin{aligned} t_{R,1} &= \sum_i (T_{1,1}(i))^2 + (C_{2,2}(i))^2; \\ t_{R,2} &= \sum_i (T_{1,2}(i))^2; \\ t_{R,3} &= \sum_i T_{1,1}(i)T_{2,2}(i); \\ t_{R,4} &= \sum_i T_{1,2}(i) (T_{1,1}(i) - T_{2,2}(i)). \end{aligned}$$

where  $\mathbf{i} = (i_3, \dots, i_R)$ . An analytical optimal value of  $\theta$  can be derived from (17). Indeed one finds after some simple algebra that for a symmetric matrix  $\mathbf{Z} = (Z_{i,j})$

$$\mathbf{u}_\theta^T \mathbf{Z} \mathbf{u}_\theta = D + E \cos(4(\theta - \alpha)) \quad (19)$$

where  $D = (Z_{1,1} + Z_{2,2})/2$  is a constant term since it does not depend on  $\theta$  and  $E$  is a non negative constant term:  $E = \sqrt{((Z_{1,1} - Z_{2,2})/2)^2 + Z_{1,2}^2}$ . The angle  $\alpha$  can be determined as

$$\alpha = \frac{1}{4} \arctan \left( Z_{1,2}, \frac{1}{2}(Z_{1,1} - Z_{2,2}) \right) \quad (20)$$

where the value of  $\arctan(y, x)$  is, by definition, the unique angle  $\beta \in (-\pi, \pi]$  for which  $\cos \beta = \frac{x}{(x^2+y^2)^{1/2}}$  and  $\sin \beta = \frac{y}{(x^2+y^2)^{1/2}}$ . Since  $D$  and  $E$  are constant and  $E$  is non negative, a maximum value (w.r.t.  $\theta$ ) denoted  $\theta_{\text{opt}}$  of the left term in (19) corresponds to the maximum value of the cosine, that is 1, yielding

$$\theta_{\text{opt}} = \alpha. \quad (21)$$

Thus one has to directly consider  $\mathbf{Z} = \mathbf{A}_R$  for the derivation of a solution. Let us remark that following the idea of Proposition 3 matrices of cumulants of different orders can be considered very easily. Indeed, let us consider matrices of cumulants of orders  $S_1, \dots, S_m$ ,  $m \in \mathbb{N}^*$  such that  $3 \leq S_1 < \dots < S_m$ . Then according to (17)

$$\sum_{i=1}^m \gamma_i \mathcal{G}_{S_i}^2(\mathbf{y}) = \mathbf{u}_\theta^T \left( \sum_{i=1}^m \gamma_i \mathbf{A}_{S_i} \right) \mathbf{u}_\theta$$

where the  $\gamma_i$ 's are real non-negative constants with at least one non zero. Hence it is sufficient to consider now the matrix  $\mathbf{Z} = \sum_{i=1}^m \gamma_i \mathbf{A}_{S_i}$  for the determination of the value of  $\theta_{\text{opt}}$  in (21), (20).

Such a generalized algorithm for joint-diagonalization of cumulant matrices is called eJADE for "extended JADE" when considering the original implementation of JADE and adding directly the new matrices to be joint-diagonalized.

## 6. COMPUTER SIMULATIONS

In order to illustrate the potential usefulness of the above results, some computer simulations are now presented. For the fourth order case, we use the original JADE algorithm in its version 1.5 of December 1997 for real signals. While for the other cases, we consider eJADE in exactly the same conditions. We use both only third order cumulant matrices whose corresponding algorithm is denoted eJADE(3), and third plus fourth order cumulant matrices whose corresponding algorithm is denoted eJADE(3,4). We first use a signal with parameterized third and fourth-order cumulants. It is a discrete i.i.d. signal called MS( $\alpha$ ) which takes its values in the set  $\{-1, 0, \alpha\}$  with the respective probability  $\{\frac{1}{1+\alpha}, \frac{\alpha-1}{\alpha}, \frac{1}{\alpha(1+\alpha)}\}$ . The real parameter  $\alpha$  called "cumulant parameter" is such that  $\alpha \geq 1$ . Hence for a MS( $\alpha$ ) signal  $a(n)$ , one easily has  $E[a] = 0$ ,  $E[a^2] = 1$ ,  $C_3[a] = \alpha - 1$  and  $C_4[a] = \alpha^2 - \alpha - 2$ . The performances of the algorithms are associated to a non negative index/measure of performance [6] which is zero when the separation holds. We have plotted both the mean and the standard deviation (STD) of the estimated index over 500 Monte Carlo runs.

**Experiment 1:** With  $N = 3$ , the two first considered sources are MS( $\alpha$ ) signals while the third one is a Gaussian i.i.d. signal. The following mixing matrix is used

$$\mathbf{G} = \begin{pmatrix} 1 & 0.9 & 0.9 \\ 0.8 & 1 & 0.9 \\ 0.8 & 0.8 & 1 \end{pmatrix}. \quad (22)$$

We plot the mean and STD of the estimated index as a function of the cumulant parameter  $\alpha$ . The data number is held constant to  $N_d = 400$ . The figure Fig.1 shows that the performances of eJADE(3,4) in comparison to JADE and eJADE(3) are less subject to variation w.r.t. the statistics of the sources.

**Experiment 2:** In this case, we consider two speech signals which are plotted in Fig.2. The components of (2,2) mixing matrix are chosen randomly with an uniform law in the interval  $[-1, 1]$ . We plot the mean and STD of the estimated index as a function of the data number taken as the first  $N_d$  samples of each speech signals. The figure Fig.3 shows that the joint-diagonalization of only third order cumulant matrices with eJADE(3) can be sufficient for the separation of speech signals with, however, lower performances. This is not surprising because the third and fourth order cumulants (estimated over the whole signals) of the speech signals we use, are respectively around 0.5 and 2.7. On the other hand, the performances of the algorithm eJADE(3,4) using both third and fourth order cumulant matrices are a little bit better.

## REFERENCES

- [1] C. Jutten and J. Herault, "Blind separation of sources, Part I: An adaptative algorithm based on neuromimetic architecture", *Signal Processing*, Vol. 24, pp 1-10, 1991.
- [2] P. Comon, "Independent component analysis, a new concept?", *Signal Processing*, Vol. 36, pp 287-314, 1994.
- [3] J.-F. Cardoso and A. Souloumiac, "Blind beamforming for non gaussian signals", *IEE Proceedings-F*, Vol. 40, pp 362-370, 1993.
- [4] E. Moreau and N. Thirion-Moreau, "Nonsymmetrical contrasts for source separation", *IEEE Trans. Signal Processing*, Vol. 47, No. 8, pp 2241-2252, August 1999.
- [5] L. De Lathauwer, *Signal processing based on multilinear algebra*, PhD Thesis, K.U. Leuven, Sept. 1997.
- [6] E. Moreau, "A generalization of joint-diagonalization criteria for source separation", Submitted to *IEEE Trans. Signal Processing*, March 1999; revised version, April 2000.

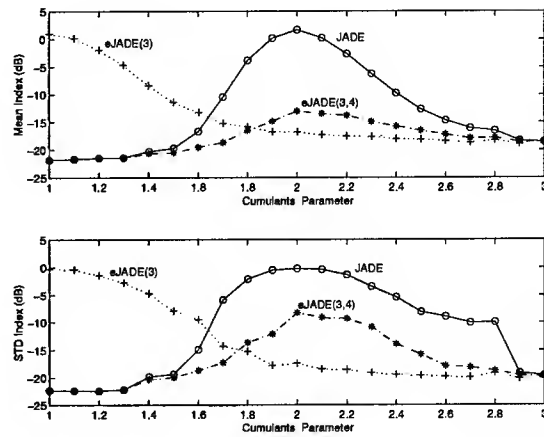


Figure 1: Mean and STD of the estimated index w.r.t. the cumulant parameter  $\alpha$ .

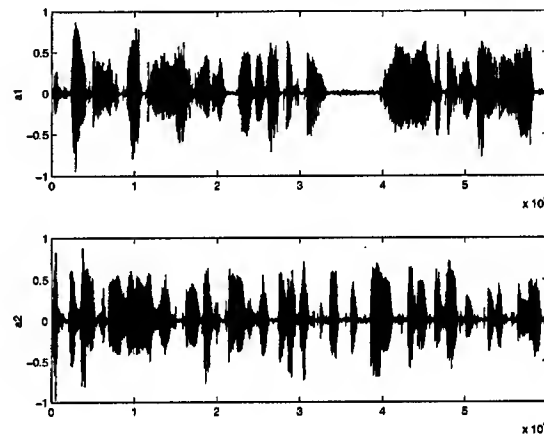


Figure 2: The two real speech signals used.

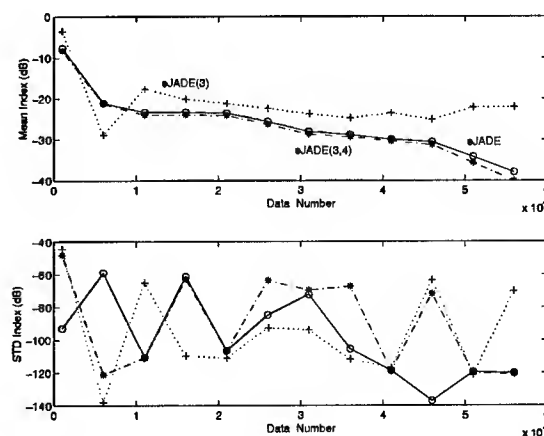


Figure 3: Mean and STD of the estimated index w.r.t. the first  $N_d$  samples of the speech signals.

# NEW CRITERIA FOR BLIND SIGNAL SEPARATION

Nadège THIRION-MOREAU and Eric MOREAU

MS-GESSY, ISITV, av. G. Pompidou, BP 56,

F-83162 La Valette du Var Cedex, France

e-mail: thirion@isitiv.univ-tln.fr, moreau@isitiv.univ-tln.fr

## ABSTRACT

The problem of multichannel blind signal deconvolution is considered. The mixing system is supposed to be stable and invertible and the input signals, also called sources, are assumed zero-mean independent and identically distributed (i.i.d) random signals. Using the hypothesis that sources are statistically independent, we propose a generalization to the convolutive case of some separation criteria available in the instantaneous one. Hence, we obtain a new generalized class of criteria for signal deconvolution.

## 1. INTRODUCTION

The problem of multichannel blind signal deconvolution (or blind equalization) of Linear Time Invariant (LTI) systems arises in various fields of engineering and applied sciences among which radio telemetry, data communication, passive radar/sonar processing, seismic exploration and so on...

In the past ten years, most of the "blindly" operating approaches have encountered a "restrictive" model commonly known as *sources separation*. Indeed, in such a problem, the coupling channels are assumed unknown yet constant gains. The goal is then to recover the inputs from the only outputs, without a priori knowledge neither about the mixing system nor about the inputs excepting their independence. In this communication, we consider the more general model implying that the coupling systems are unknown LTI systems. Some solutions based on high order statistics have been proposed recently in the literature [8]-[14]. Here, we focus on those which are based on the optimization of contrasts functions or contrasts (previously developed in the context of sources separation) involving high order statistics.

After some recalling on contrasts, we demonstrate that it is possible to consider contrasts involving cross-cumulants (of order strictly higher than two), in the convolutive case. New contrasts of this type are then presented.

## 2. PROBLEM FORMULATION

We consider the multichannel LTI and generally non-causal system described by the following equation:

$$\mathbf{x}(n) = \sum_k \mathbf{G}(k) \mathbf{a}(n-k) \quad (1)$$

where  $n \in \mathbb{Z}$  is the time index,  $\mathbf{a}(n)$  is the  $(N, 1)$  vector of *statistically independent* sources,  $\mathbf{x}(n)$  is the  $(N, 1)$  vector of observations and

$$\{\mathbf{G}\} \triangleq \{\mathbf{G}(n), n \in \mathbb{Z}\}$$

is a sequence of  $(N, N)$  matrices which describe the impulse response of the LTI mixing filter. The  $(N, N)$  transfer matrix of system  $\{\mathbf{G}\}$  is also introduced:

$$[\hat{\mathbf{G}}] \stackrel{\text{def}}{=} \hat{\mathbf{G}}(z) \stackrel{\text{def}}{=} \sum_k \mathbf{G}(k) z^{-k} \quad (2)$$

where  $z^{-1}$  stands for the time-delay operator.

The multichannel blind deconvolution problem consists in estimating a LTI filter (equalizer)  $\{\mathbf{H}(\cdot)\}$  thanks to the only outputs (observations)  $\mathbf{x}(n)$  of an unknown LTI system  $\{\mathbf{G}\}$  and such that the vector:

$$\mathbf{y}(n) = \sum_k \mathbf{H}(k) \mathbf{x}(n-k) \quad (3)$$

restores the  $N$  input signals  $a_i(n)$ ,  $i = 1, \dots, N$ .

In this context, it is useful to define the global LTI filter  $\{\mathbf{S}(\cdot)\}$  according to

$$\mathbf{y}(n) = \sum_k \mathbf{S}(k) \mathbf{a}(n-k) \quad (4)$$

whose transfer function is

$$\hat{\mathbf{S}}(z) = \sum_k \mathbf{S}(k) z^{-k}.$$

The global system is illustrated in Fig.1.

To solve this problem, we make the following assumptions:

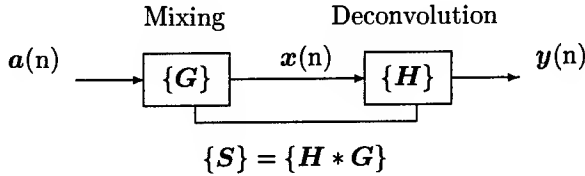


Figure 1: The global system.

A1. Each input  $a_i(n)$ ,  $i = 1, \dots, N$  is a zero-mean independent and identically distributed (i.i.d.) discrete random signal. Without any loss of generality  $a_i(n)$ ,  $i = 1, \dots, N$  can be assumed unit power. Moreover, we will assume that the cumulants of the random sources exist and that for a given order  $R$ , at most one of them has a null cumulant. We recall that the  $R$ -th order cumulant ( $R \in \mathbb{N}^*$ ) is defined as

$$\text{Cum}[a_i(n), \dots, a_i(n)]_{R \times}$$

For a i.i.d. signal, it does not depend on time any more. So, it is simply noted  $C_R[a_i]$ .

A2. The unknown LTI system  $\{G(\cdot)\}$  is assumed stable and invertible.

As sources are assumed inobservable, some inherent indeterminations subsist when they are restored. In fact, in general, the order, the power and the time origin of each components of the source vector  $a(n)$  can not be recovered. Indeed, the multichannel blind deconvolution problem combines the inherent indeterminations of the source separation problem together with the inherent indeterminations of the classical blind scalar deconvolution problem. That is why, we consider that the multichannel blind deconvolution problem is solved if and only if (iff) the global LTI system  $\{S(\cdot)\}$  reads:

$$\begin{aligned} S(z) &\stackrel{\text{def}}{=} \sum_k S(k)z^{-k} \\ &= D(z)D_1P \end{aligned} \quad (5)$$

where  $D(z)$  is a diagonal matrix such that its entries are

$$d_{ii}(z) = z^{-n_i}, \quad i = 1, \dots, N$$

$n_i$  are integers,  $D_1$  is an invertible constant diagonal matrix and  $P$  is a permutation matrix. Finally the global system is assumed to be

A3. "paraunitary" i.e.  $\hat{S}(z)$  satisfies

$$\hat{S}(z)\hat{S}^T(z^{-1}) = I$$

where  $I$  denotes the  $(N, N)$  identity matrix.

It can be noticed that this last assumption can be made without loss of generality since it is equivalent to guarantee that the signal  $y(n)$  is white, i.e.

$$E[y(n)y^T(n-k)] = I\delta[k]$$

where  $E[\cdot]$  denotes the mathematical expectation and  $\delta[k] = 1$  if  $k = 0$  and 0 otherwise. This constraint can always be satisfied provided that a classical prewhitening of the observations is performed.

### 3. CONTRAST FUNCTIONS

First introduced for instantaneous mixing [2], contrast functions have been recently generalized to the convolutive case [3, 8, 10]. Before recalling the definition of contrasts, let us first introduce some useful notations:  $\mathcal{A}$  stands for the set of random vectors that satisfy hypothesis A1.  $\mathcal{S}$  stands for the set of systems  $\hat{S}(z)$  satisfying hypothesis A2 and A3. The subset of  $\mathcal{S}$  of systems conforming (5) is denoted by  $\mathcal{P}$ . The set of random vectors  $y(n)$  built from (4) with  $a(n) \in \mathcal{A}$  and  $\hat{S}(z) \in \mathcal{S}$  is denoted by  $\mathcal{Y}_A$ .

Let us now recall the more general definition of contrasts one has to consider in the convolutive case [10]:

**Definition 1** A contrast on  $\mathcal{Y}_A$  is a multivariable function  $\mathcal{I}(\cdot)$  mapping  $\mathcal{Y}_A$  on  $\mathbb{R}$ , and satisfying the three following requirements:

- P1.  $\forall y \in \mathcal{Y}_A, \forall \hat{D}(z) \in \mathcal{D}, \mathcal{I}([\hat{D}]y) = \mathcal{I}(y)$  ;
- P2.  $\forall a \in \mathcal{A}, \forall \hat{S}(z) \in \mathcal{S}, \mathcal{I}([\hat{S}]a) \leq \mathcal{I}(a)$  ;
- P3.  $\forall a \in \mathcal{A}, \forall \hat{S}(z) \in \mathcal{S}, \mathcal{I}([\hat{S}]a) = \mathcal{I}(a) \Rightarrow \hat{S}(z) \in \mathcal{P}$ .

One of the first contrasts available in the convolutive case has been exhibited by P. Comon [3]. It reads:

$$\mathcal{I}_R(y) = \sum_{i=1}^N (C_R[y_{i1}])^2 \quad (6)$$

where  $R$  is greater than or equal to three. It is a direct extension of a contrast previously proposed in the instantaneous case [2].

It is also interesting to keep in mind the two following properties of contrasts [10]:

**Property 1** Given  $\mathcal{I}_1(\cdot)$  a function of  $\mathcal{Y}_A$  on  $\mathbb{R}$  and  $\mathcal{I}_2(\cdot)$  a contrast on  $\mathcal{Y}_A$ . If

$$\begin{aligned} \forall y \in \mathcal{Y}_A, \quad \mathcal{I}_1(y) &\leq \mathcal{I}_2(y) \\ \forall a \in \mathcal{A}, \quad \mathcal{I}_1(a) &= \mathcal{I}_2(a) \end{aligned} \quad (7)$$

then  $\mathcal{I}_1(\cdot)$  is a contrast on  $\mathcal{Y}_A$ .

The second one is:

**Property 2** If  $\mathcal{I}_1(\cdot)$  is a contrast on  $\mathcal{Y}_A$  then  $\forall \alpha \in \mathbb{R}_+^*$  and  $\forall \beta \in \mathbb{R}$ ,  $\mathcal{I}_2(\cdot) = \alpha \mathcal{I}_1(\cdot) + \beta$  is also a contrast on  $\mathcal{Y}_A$ .

This last property allows us to define the notion of equivalent contrast as

**Definition 2** If  $\mathcal{I}_1(\cdot)$  and  $\mathcal{I}_2(\cdot)$  are two contrasts in the sense of Property 2 then they are said to be equivalent.

## 4. NEW AND GENERALIZED RESULTS

### 4.1. Contrasts with cross-cumulants

Our main goal is to generalize some contrasts available in the instantaneous case to the convolutive one. At this aim, let us introduce the following notation

$$\mathcal{C}_R^y[i, \ell] = \text{Cum}[y_{i_1}(n), y_{i_1}(n), y_{i_2}(n - \ell_2), \dots, y_{i_{R_1}}(n - \ell_{R_1})] \quad (8)$$

where  $R$  stands for the cumulant order,  $R_1 = R - 1$  and

$$\begin{aligned} i &= (i_1, \dots, i_{R_1}) \\ \ell &= (\ell_2, \dots, \ell_{R_1}) . \end{aligned}$$

We have the following first result:

**Proposition 1** Let  $R$  be an integer such that  $R \geq 3$ , the function

$$\mathcal{J}_{i,R}(\mathbf{y}) = \sum_{i, \ell} (\mathcal{C}_R^y[i, \ell])^2 \quad (9)$$

is a contrast for white vector  $\mathbf{y}(n)$ .

*Proof.* To demonstrate this result, let us begin with some important second order results. Because of assumption A3, the inverse of the global system is  $\hat{\mathbf{S}}^T(1/z)$  and then

$$\mathbf{a}(n) = \sum_k \mathbf{S}^T(-k) \mathbf{y}(n - k)$$

that is component wise

$$a_j(n) = \sum_{l, i} S_{ij}(-l) y_i(n - l) .$$

Consequently, denoting

$$R_{a_{j_1}, a_{j_2}}(k_1, k_2) = \mathbb{E}[a_{j_1}(n + k_1) a_{j_2}(n + k_2)]$$

and using the fact that  $\mathbf{y}(n)$  is a white vector, we have

$$\begin{aligned} R_{a_{j_1}, a_{j_2}}(k_1, k_2) &= \sum_{l_1, l_2} \sum_{i_1, i_2} S_{i_1 j_1}(-l_1) S_{i_2 j_2}(-l_2) \\ &\quad \mathbb{E}[y_{i_1}(n + k_1 - l_1) y_{i_2}(n + k_2 - l_2)] \\ &= \sum_{l_1, i_1} S_{i_1 j_1}(k_2 - l_1) S_{i_1 j_2}(k_1 - l_1) \end{aligned}$$

On the other hand, we also have

$$R_{a_{j_1}, a_{j_2}}(k_1, k_2) = \delta[j_1 - j_2] \delta[k_1 - k_2]$$

Then

$$\sum_{l_1, i_1} S_{i_1 j_1}(k_2 - l_1) S_{i_1 j_2}(k_1 - l_1) = \delta[j_1 - j_2] \delta[k_1 - k_2] \quad (10)$$

and thus

$$\sum_{k_1, i_1} (S_{i_1 j_1}(k_1))^2 = 1 . \quad (11)$$

Involving that

$$\sum_{k_1, i_1} (S_{i_1 j_1}(k_1))^4 \leq \sum_{k_1, i_1} (S_{i_1 j_1}(k_1))^2 = 1 .$$

Using the multilinearity property of cumulants and using (10), we have

$$\mathcal{J}_{i,R}(\mathbf{y}) = \sum_{j_1, k_1} \left( \sum_{i_1} (S_{i_1 j_1}(k_1))^4 \right) (\mathcal{C}_R[a_{j_1}])^2$$

and then

$$\mathcal{J}_{i,R}(\mathbf{y}) \leq \sum_{j_1} (\mathcal{C}_R[a_{j_1}])^2 = \mathcal{J}_{i,R}(\mathbf{a})$$

Moreover it is easy to see that we have the equality if and only if  $\hat{\mathbf{S}}(z)$  satisfies (5), yielding that  $\mathcal{J}_{i,R}(\mathbf{y})$  is a contrast.  $\diamond$

This result can be seen as a first generalization to the convolutive case of the underlying contrast of JADE algorithm given in [1]. But  $R - 2$  delays have to be taken into account in the contrast which can make its optimization rather cumbersome. Let us thus simplify  $\mathcal{J}_{i,R}(\mathbf{y})$ . At this aim, we now consider the following parameterized vector of delays

$$\ell_\alpha = (\ell_{\alpha(2)}, \dots, \ell_{\alpha(R_1)})$$

where  $\alpha(\cdot)$  is any application from the set  $\{2, \dots, R_1\}$  to the set  $\{0, 2, \dots, R_1\}$ . Moreover, by definition, if  $\alpha(i) = 0$  for a given  $i \in \{2, \dots, R_1\}$  then no corresponding delay is considered. Using the notation (8) with  $\ell_\alpha$ , we have the following result:

**Proposition 2** Let  $R$  be an integer such that  $R \geq 3$ , the function

$$\mathcal{J}_R^\alpha(\mathbf{y}) = \sum_{i, \ell_\alpha} (\mathcal{C}_R^\alpha[i, \ell_\alpha])^2 \quad (12)$$

is a contrast for white vector  $\mathbf{y}(n)$ .

*Proof.* We have the two following relations

$$\begin{aligned} \mathcal{J}_R^\alpha(\mathbf{y}) &\leq \mathcal{J}_{i,R}(\mathbf{y}) \\ \mathcal{J}_R^\alpha(\mathbf{a}) &= \mathcal{J}_{i,R}(\mathbf{a}) \end{aligned} \quad (13)$$

Using the result in Proposition 1, the Property 1 allows us to conclude that  $\mathcal{J}_R^\alpha(\mathbf{y})$  is a contrast.  $\diamond$

In order to illustrate this proposition, we present now some examples. First, if  $\alpha(\cdot)$  is the identity application, i.e.  $\alpha(\cdot) = \text{Id}(\cdot)$ , then

$$\mathcal{J}_R^{\text{Id}}(\mathbf{y}) = \mathcal{J}_{i,R}(\mathbf{y}) .$$

Hence Proposition 2 is a generalization of Proposition 1. Now if  $\alpha(\cdot)$  is the null application, i.e.  $\alpha(\cdot) = \mathbf{O}(\cdot)$  where  $\forall k, \mathbf{O}(k) = 0$ , then no delay is taken into consideration and the contrast simply reads

$$\mathcal{J}_R^\mathbf{O}(\mathbf{y}) = \sum_i (\text{Cum}[y_{i_1}, y_{i_1}, y_{i_2}, \dots, y_{i_{R_1}}])^2 .$$

This result is now the direct generalization of the underlying contrast of the JADE algorithm to the convolutive case. This is the only case where no delays have to be considered. Thus this latter contrast contains the minimum number of cumulants. For fourth order cumulants, if  $\alpha(1) = 0$  and  $\alpha(2) = 2$  and denoting  $\alpha_1(\cdot)$  this application, we have the following contrast

$$\mathcal{J}_R^{\alpha_1}(\mathbf{y}) = \sum_{i, \ell_2} (\text{Cum}[y_{i_1}(n), y_{i_1}(n), y_{i_2}(n), y_{i_3}(n - \ell_2)])^2$$

On the other hand, if  $\alpha(1) = 1$  and  $\alpha(2) = 1$  and denoting  $\alpha_2(\cdot)$  this application, we have the following contrast

$$\mathcal{J}_R^{\alpha_2}(\mathbf{y}) = \sum_{i, \ell_1} (\text{Cum}[y_{i_1}(n), y_{i_1}(n), y_{i_2}(n - \ell_1), y_{i_3}(n - \ell_1)])^2$$

This is just few examples of the family of contrasts one can find using Proposition 2.

#### 4.2. Contrasts with parameterized cross-cumulants

Now, to simplify, let us focus on the case of fourth-order cumulants with no delays in the contrasts. We define then the following three functions

$$\mathcal{C}_1(\mathbf{y}) = \sum_{\substack{i_1, i_2=1 \\ i_2 \neq i_1}}^N (\text{Cum}[y_{i_1}, y_{i_1}, y_{i_1}, y_{i_2}])^2 ;$$

$$\mathcal{C}_2(\mathbf{y}) = \sum_{\substack{i_1, i_2=1 \\ i_2 > i_1}}^N (\text{Cum}[y_{i_1}, y_{i_1}, y_{i_2}, y_{i_2}])^2 ;$$

$$\mathcal{C}_3(\mathbf{y}) = \sum_{\substack{i_1, i_2, i_3=1 \\ i_3 \neq i_2 \neq i_1 \\ i_3 > i_2}}^N (\text{Cum}[y_{i_1}, y_{i_1}, y_{i_2}, y_{i_3}])^2 .$$

These functions contain only cross-cumulants of different types. Then we have the following result:

**Proposition 3** Consider three real numbers  $\alpha_i, i = 1, \dots, 3$  such that  $\forall i, \alpha_i \leq 1$ , then, denoting  $\alpha = (\alpha_1, \alpha_2, \alpha_3)$  the function

$$\mathcal{J}_{4,\alpha}(\mathbf{y}) = \mathcal{I}_4(\mathbf{y}) + 2(\alpha_1 \mathcal{C}_1(\mathbf{y}) + \alpha_2 \mathcal{C}_2(\mathbf{y}) + \alpha_3 \mathcal{C}_3(\mathbf{y})) \quad (14)$$

is a contrast for white vectors  $\mathbf{y}(n)$ .

*Proof.* Because  $\forall i, \alpha_i \leq 1$ , we have the two following relations

$$\begin{aligned} \mathcal{J}_{4,\alpha}(\mathbf{y}) &\leq \mathcal{J}_4^\mathbf{O}(\mathbf{y}) \\ \mathcal{J}_{4,\alpha}(\mathbf{a}) &= \mathcal{J}_4^\mathbf{O}(\mathbf{a}) \end{aligned} \quad (15)$$

Recalling that  $\mathcal{J}_4^\mathbf{O}(\mathbf{y})$  is a contrast, the property 1 allows us to conclude that  $\mathcal{J}_{4,\alpha}(\mathbf{y})$  is a contrast.  $\diamond$

Let us first notice that if  $\alpha = \alpha_o$  where  $\alpha_o = (0, 0, 0)$  then

$$\mathcal{J}_{4,\alpha_o}(\mathbf{y}) = \mathcal{I}_4(\mathbf{y})$$

which is the contrast in (6). On the other hand, if  $\alpha = \alpha_1$  where  $\alpha_1 = (1, 1, 1)$  then

$$\mathcal{J}_{4,\alpha_1}(\mathbf{y}) = \mathcal{J}_4^\mathbf{O}(\mathbf{y}) .$$

All other values of  $\alpha_i, i = 1, \dots, 3$ , yield a new contrast.

For simplicity, we have restricted our attention to the case of fourth order cumulants with no delays in the cross-cumulants. But using the results in Proposition 2, Proposition 3 can be easily generalized to any order of cumulants with or without delays.

More generally, if a given contrast involves cross-cumulants, then following the same principle as the one applied to find the above result, one can parameterize alike its cross-cumulants.

## 5. CONCLUSION

In this paper, we have generalized some contrasts available in the case of instantaneous mixtures to the convolutive one. We have shown that cross-cumulants can be used in this case and that they can be parameterized which leads to a more general family of contrasts. In all cases, delays can be taken into consideration in order to use temporal statistical informations.

## REFERENCES

- [1] J.F. Cardoso and A. Souloumiac, "Blind beam-forming for non gaussian signals", *IEE Proceedings F*, Vol. 40, pp 362-370, 1993.
- [2] P. Comon, "Independent component analysis, a new concept?", *Signal Processing*, Vol. 36, pp 287-314, 1994.
- [3] P. Comon, "Contrasts for multichannel blind deconvolution", *IEEE Signal Processing Letters*, Vol. 3, No. 7, pp 209-211, July 1996.
- [4] K.I. Diamantaras, A.P. Petropulu and B. Chen, "Blind two-input two-output FIR channel identification based on frequency domain second-order statistics", *IEEE Trans. Signal Processing*, Vol. 48, No. 2, pp 534-542, Feb. 2000.
- [5] Y. Inouye and K. Hirano, "Cumulant-based blind identification of linear multi-input multi-output systems driven by colored inputs", *IEEE Trans. Signal Processing*, Vol. 45, No. 6, pp 1543-1552, June 1997.
- [6] P. Loubaton and P. Regalia, "Blind deconvolution of multivariate signals: a deflation approach", in Proc. *ICC'93, Int. Conf. on Communication*, Geneva, Switzerland, Vol. 2, pp 1160-1164, May 1993.
- [7] E. Moreau, "A block algorithm for blind signal deconvolution", in Proc. *SPAWC'97, IEEE Signal Processing Workshop on Signal Processing Advances in Wireless Communications*, Paris, France, pp 93-96, April 1997.
- [8] E. Moreau and N. Thirion, "Multichannel blind signal deconvolution using high order statistics", in Proc. *SSAP'96*, Corfu, Greece, pp 336-339, June 1996.
- [9] E. Moreau and J.-C. Pesquet, "Generalized Contrasts for Multichannel Blind Deconvolution of Linear Systems", *IEEE Signal Processing Letters*, Vol. 4, No. 6, pp 182-183, June 1997.
- [10] E. Moreau, J.-C. Pesquet and N. Thirion-Moreau, "An equivalence between non symmetrical contrasts and cumulant matching for blind signal separation", in Proc. *First Int. Workshop on Independent Component Analysis and Signal Separation, (ICA'99)*, Aussois, France, pp 301-306, Jan. 1999.
- [11] C. Simon, P. Loubaton, C. Vignat, C. Jutten and G. d'Urso, "Separation of a class of convolutive mixtures: a contrast function approach", in Proc. *ICASSP'99*, Phoenix, Arizona, USA, May 1999.
- [12] A. Swami, G. Giannakis, and S. Shamsunder, "Multichannel ARMA processes", *IEEE Trans. Signal Processing*, Vol. 42, No. 4, pp 898-913, April 1994.
- [13] J.K. Tugnait, "On blind separation of convolutive mixtures of independent linear signals", in Proc. *SSAP'96, IEEE SP Workshop on SSAP*, Corfu, Greece, pp 312-315, June 1996.
- [14] J. K. Tugnait, "On blind separation of convolutive mixtures of independent linear signals in unknown additive noise", *IEEE Trans. Signal Processing*, Vol. 46, No. 11, pp 3117-3123, Nov. 1998.
- [15] D. Yellin and E. Weinstein, "Criteria for multichannel signal separation", *IEEE Trans. Signal Processing*, Vol. 42, No. 8, pp 2158-2168, August 1994.



# AN ITERATIVE ALGORITHM USING SECOND ORDER MOMENTS APPLIED TO BLIND SEPARATION OF SOURCES WITH SAME SPECTRAL DENSITIES

*Jean-François Cavassilas, Bernard Xerri and Bruno Borloz*

MS/GESSY - Université de Toulon et du Var  
Avenue Georges Pompidou, BP 56  
83162 La Valette du Var Cedex (FRANCE)  
Fax: 33 494 142 598  
xerri@isitv.univ-tln.fr  
borloz@univ-tln.fr

## ABSTRACT

In this paper, we are interested in the separation of  $N$  independent sources recorded simultaneously by  $N$  receivers. The mixture is realized instantaneously through an unknown constant matrix  $M$ .

When the spectral densities of the sources are different, several methods using second order moments have been proposed whose results are convincing. Nevertheless, these methods are no more efficient when their spectral densities are the same. Our talk is interested in this special case where sources may even be white. The method we propose is based on the evaluation of second order moments estimated from extracted series of the observations. We will talk of conditional second order moments.

An iterative algorithm is proposed which calculates, at each step, a matrix  $K_i$  so that  $K_n K_{n-1} \dots K_1 M$  tends, when  $n$  increases, to  $D\Pi$ , product of a diagonal matrix and a permutation matrix.

We show that restrictive conditions on the probability distributions of the sources must be verified to assure the separation.

In the two-dimensional case, we prove that the algorithm separates uniformly distributed sources, and that it doesn't separate gaussian sources.

The algorithm proposed is robust towards the number of sources; simulations with more than 20 uniformly distributed sources were successful.

## 1. INTRODUCTION

We present a new iterative method to separate  $N$  independent sources instantaneously mixed through an unknown constant matrix  $M$ . It is important to note that the spectral densities of the sources may be the

same.

First we introduce the mixture model. Second we explain the method used to retrieve the sources and the restrictive conditions necessary to reach the separation. Then we restrict our talk to the two dimensional mixing case where calculation are easier to do. In this particular case, conditions of separation are given. The case of uniformly distributed sources is treated with more details. The gaussian case is proved to be impossible to separate by such a method.

Finally, results obtained on simulated data are shown.

## 2. BASIC ASSUMPTIONS AND MODEL

Let consider  $N$  sources assumed to be centered, stationary and statistically independent. These sources can be represented by a matrix  $X = [x_1 \ x_2 \ \dots \ x_N]^t$  of dimension  $N \times L$ . They are simultaneously recorded by a set of  $N$  receivers: the available observations are represented by a matrix  $Y = [y_1 \ y_2 \ \dots \ y_N]^t$  of dimension  $N \times L$ .

The observations  $Y$  are linked to the sources  $X$  by the linear relationship:

$$Y = MX, \quad (1)$$

where  $M$  is an unknown constant matrix.

What's more, for reasons explained later, we will restrict our topic to identically distributed sources whose probability distributions are symmetrical.

Note that vectors are noted in bold font.

## 3. THE IDEA OF THIS ALGORITHM

The now classical methods using second order statistics (AMUSE [1] [2], SOBI [3] or IMISO [4]) to separate

instantaneous mixtures of independent sources proved they were efficient when the power spectral densities (PSD) of the sources are different. For example, the IMISO algorithm exploits the covariance matrix and its second derivative: it has been proved that if the PSD are not identical, conditions are gathered to succeed. However, if these conditions are not verified, they fail.

In the case of sources with same PSD, it is no more possible to calculate a matrix  $C$  such as  $CY = D\Pi X$  (we retrieve  $X$  except for a diagonal matrix  $D$  and a permutation matrix  $\Pi$ ), and the way of an iterative formulation seemed attractive.

The idea is always to find a linear transformation that makes independent the observations, but using conditional and recursive methods.

#### 4. PRESENTATION OF THE RECURSIVE ALGORITHM

Historically, this method was first used to separate two white uniformly distributed sources. But the algorithm had to be modified to take the specific aspects of larger problems into account. So, this section is broken down in two parts to reproduce this approach.

##### 4.1. General topic

The results described in this subsection are general. From the observations  $Y$ , let extract a sequence of the first observation  $y_1(t)$ : we select the indices  $t = 1$  to  $L$  for which  $y_1(t)$  is positive (*i.e.* we create a vector  $S$  whose elements are these indices; the length of  $S$  is  $L_s$ ). Then we create extracted observations  $Y_S$  ( $Y_S$  is a matrix of dimensions  $N \times L_s$ ). That means that for  $i = 1$  to  $N$  and  $t = 1$  to  $L_s$ , the  $i^{th}$  component of  $Y_S$  is the vector  $y_{S_i}(t) = y_i(S(t))$  of dimension  $L_s$ .

With the constraint chosen here,  $E\{Y_S\} = \tilde{Y}_S \neq 0$  and we will note  $\tilde{Y}_S = Y_S - \tilde{Y}_S$ .

From these extracted observations we calculate the covariance matrices of  $Y_S$  and  $\tilde{Y}_S$  noted:

$$N_1 = E\{Y_S Y_S^t\}$$

and

$$N_2 = E\{\tilde{Y}_S \tilde{Y}_S^t\}$$

which are  $N \times N$  matrices.

We will use the same notation for  $X_S$  the extracted sequence of  $X$  corresponding to  $S$ ,  $E\{X_S\} = \tilde{X}_S = m$  and  $\tilde{X}_S = X_S - m$ .

From equation (1) we can deduct:

$$N_1 = ME\{X_S X_S^t\} M^t,$$

$$N_2 = ME\{\tilde{X}_S \tilde{X}_S^t\} M^t = M[E\{X_S X_S^t\} - mm^t] M^t.$$

Let note  $E\{X_S X_S^t\} = D_S$ . In the general case,  $D_S$  is not diagonal; however, if the laws of  $X_i$  are symmetrical,  $D_S$  is diagonal. We will restrict our talk, to simplify, to the case of sources  $X_i$  with the same symmetrical density distribution; this assumption assures that  $D_S$  becomes a scalar matrix ( $D_S = \sigma_S \mathbb{I}$  where  $\mathbb{I}$  is the identity matrix).

Now let define:

$$\Gamma = N_1^{-1} N_2 = M^{-t} [\mathbb{I} - D_S^{-1} mm^t] M^t,$$

and

$$G = \mathbb{I} - D_S^{-1} mm^t.$$

Note that neither  $X_S$  nor  $m$ ,  $D_S$  and then  $G$  are directly attainable.

Let denote  $P$  and  $Q$  the eigenvectors and eigenvalues of  $\Gamma = M^{-t} G M^t$ , *i.e.* such as  $\Gamma P = P Q$ .

Then  $G M^t P = M^t P Q$ .

If we note  $V = M^t P$ ,  $V$  and  $Q$  are the eigenvectors and eigenvalues of  $G$  (*i.e.*  $GV = VQ$ ).

$P$  and  $Q$  depend on  $Y$  and are then calculable, while  $V$  is not because it depends on  $M$ .

In the particular case of sources with the same symmetrical density distribution

$$\Gamma = M^{-t} [\mathbb{I} - \sigma_S^{-1} mm^t] M^t. \quad (2)$$

We know that the eigenvectors of  $mm^t$  are  $m$  and  $m_j^\perp$  for  $j = 1$  to  $N-1$  with respectively the eigenvalue  $m^t m$  (multiplicity 1) and 0 (multiplicity  $N-1$ ), where the  $m_j^\perp$  are the  $N-1$  vectors orthogonal to  $m$ .

Thus, the eigenvectors of  $G$  are  $m$  (with the eigenvalue  $1 - \sigma_S^{-1} m^t m$ ) and the  $m_j^\perp$  (with the eigenvalue 1).

Let's write the equation (1) as:

$$Y^{(1)} = M^{(1)} X, \quad (3)$$

( $Y^{(1)} = Y$  and  $M^{(1)} = M$ ) that denotes the first iteration of our algorithm.

Applying  $P^t$  to the initial observations  $Y^{(1)}$ , we obtain a combination of the  $y_i$  noted  $Y^{(2)} = P^t Y^{(1)}$  called the "new observations". Then,

$$Y^{(2)} = P^t Y^{(1)} = P^t M^{(1)} X = M^{(2)} X$$

appears as a new mixture of sources,  $M^{(2)}$  being the mixing matrix:

$$M^{(2)} = P^t M^{(1)} = V^t, \quad (4)$$

The algorithm will be efficient if  $M^{(2)}$  is closer than  $M^{(1)}$  to a diagonal matrix (except for a permutation matrix). So the properties of  $M^{(2)}$  must be studied to prove the advantages of the method.

Here, the first step of our algorithm is finished. If properties of  $M^{(2)}$  are satisfying, this step can be iterated again on the new observations  $Y^{(2)}$ .

#### Summarized Basic Algorithm

- 1) Find the indexes for which the first observation is positive and put them in a vector  $\mathbf{S}$
- 2) Create the extracted observations  $Y_S$
- 3) Compute the centered extracted observation  $\tilde{Y}_S$
- 4) Estimate the observation covariance matrices  $N_0 = E\{\tilde{Y}_S \tilde{Y}_S^t\}$  and  $N_1 = E\{Y_S Y_S^t\}$ .
- 5) Compute  $\Gamma = N_1^{-1} N_0$  and its eigenvectors  $P$ .
- 6) Compute the transformed observations  $P^t Y$ .

### 4.2. Estimation of the performance of the algorithm

To quantify the performances reached, we use the now well accepted performance criterion, positive real value which permits to know how far is a matrix  $M$  from the product  $D\Pi$  of a diagonal and a permutation matrix. This value, noted  $ind(M)$  is zero if  $M = D\Pi$  and is at worst equal to  $N$  the dimension of  $M$ .

The stopping of the iterations is done when  $ind(P)$  is smaller than a value defined in advance.

### 4.3. Limitation of the method

The proof of the efficiency of the algorithm must be given for each particular sort of probability distribution. The theoretical calculation of  $V$  (and therefore of  $m_1$  and  $m_2$ ) depends on the probability distribution of the sources, and may hardly be done in a general way. For all these reasons we will focus our attention on two remarkable cases of particular interest: the gaussian and uniform ones.

### 4.4. 2D separation problem : $N = 2$

In this particular case, we note  $\mathbf{m} = [m_1 \ m_2]^t$  and  $\mathbf{m}^\perp = [-m_2 \ m_1]^t$ .

We will note  $M$  as follow:

$$M = \begin{bmatrix} a & b \\ c & d \end{bmatrix},$$

and without loss of generality we will suppose that  $a \neq b$ , and even  $|a| > |b|$  (of course,  $M$  is invertible).

The normalized matrix  $V = M^t P$  verifies

$$V = \frac{1}{\sqrt{m_1^2 + m_2^2}} \begin{bmatrix} m_1 & -m_2 \\ m_2 & m_1 \end{bmatrix} \quad (5)$$

except for a permutation matrix; the eigenvalues matrix is given by

$$Q = \begin{bmatrix} 1 - \frac{m_1^2 + m_2^2}{\sigma_s^2} & 0 \\ 0 & 1 \end{bmatrix}. \quad (6)$$

#### 4.4.1. Whitened observations

If observations are whitened,  $Y$  become  $Y_w$  which verifies :  $Y_w = M_w X$  where  $M_w$  can be written

$$M_w = \begin{bmatrix} \alpha & \beta \\ -\beta & \alpha \end{bmatrix}.$$

This case is interesting because the performance index of  $M_w$  is easily calculated:  $ind(M_w) = 2 \left| \frac{\min(|\alpha|, |\beta|)}{\max(|\alpha|, |\beta|)} \right|^2$ . We can deduct  $P = k M^{-t} V$  from (5), where  $k$  is a constant used to normalize  $P$ :

$$P = \frac{k}{(\alpha^2 + \beta^2) \sqrt{m_1^2 + m_2^2}} \begin{bmatrix} \alpha & \beta \\ -\beta & \alpha \end{bmatrix} \begin{bmatrix} m_1 & -m_2 \\ m_2 & m_1 \end{bmatrix}.$$

The normalization of  $P$  imposes the value of  $k$  and implies  $P = \sqrt{\alpha^2 + \beta^2} M^{-t} V$ .

Then,  $M^{(2)}$  becomes

$$M^{(2)} = \sqrt{\frac{\alpha^2 + \beta^2}{m_1^2 + m_2^2}} \begin{bmatrix} m_1 & m_2 \\ -m_2 & m_1 \end{bmatrix}. \quad (7)$$

The algorithm is efficient if  $ind(M^{(2)})$  is smaller than  $ind(M^{(1)})$ , that means

$$\left| \frac{\min(|m_1|, |m_2|)}{\max(|m_1|, |m_2|)} \right| < \left| \frac{\min(|\alpha|, |\beta|)}{\max(|\alpha|, |\beta|)} \right|. \quad (8)$$

There still remains the problem of calculating  $m_1$  and  $m_2$ . As mentioned previously, the result depends only of :

→ the probability distributions of the sources,

→ the parameters  $\alpha$  and  $\beta$ .

#### 4.4.2. Not whitened observations

In that case, calculation are almost the same. The difference lies in the fact that  $V$  is found except for the power; that implies that  $V$  is multiplied to the right by a diagonal matrix  $\Lambda = \text{diag}\{\Lambda_i\}_{i,j=1,\dots,N}$ . Then,

$M^{(2)} = V^t$  is in fact equal to  $\Lambda^t V^t$ ; each line of  $M^{(2)}$  is multiplied by the constant  $\Lambda_i$ :

$$M^{(2)} = \begin{bmatrix} \Lambda_1 m_1 & -\Lambda_1 m_2 \\ \Lambda_2 m_2 & \Lambda_2 m_1 \end{bmatrix}.$$

Nevertheless, as sources can be retrieved only except for their power, the performance index of  $M^{(2)}$  is the same than in the previous subsection.

#### 4.5. $N > 2$ problem

If  $N > 2$ , this algorithm needs to be adapted. As seen before, there is indeed only one eigenvector whose eigenvalue of multiplicity one. It appears experimentally that the algorithm applied strictly as described above leads to one of the sources. In the particular case  $N = 2$  however, when one source is found, the second one is automatically found too.

But, using another observation to find  $\mathbf{S}$  will lead to another source, and so on ... Practically, the main difficulty lies in the fact of knowing what observation has converged to a source.

Several ideas were exploited :

- Whitening the observations  $Y$ , we know that at each step, the transformation matrix  $P$  must be orthogonal. The algorithm is modified as follows : the steps of the previous algorithm are performed successively for each of the  $N-1$  first observations; each time we retain the eigenvector  $\mathbf{p}_i$  of  $P$  which corresponds to the non multiple eigenvalue of  $\Gamma$ . We create a matrix  $P^\sharp$  from these  $N-1$  orthogonal vectors, the  $N^{th}$  is created from the  $N-1$  first ones.  $P^\sharp$  is orthogonal. Experimentally, this modified algorithm converges to the  $N$  sources.

As mentioned above, there remains the difficulty of the recognition of the sources among the signals obtained.

- Creating  $\mathbf{S}_i$  and  $\mathbf{S}_j$  corresponding to the observations  $\mathbf{y}_i$  and  $\mathbf{y}_j$ , we calculate the following matrices

$$N_1 = E\{\tilde{Y}_{S_i} \tilde{Y}_{S_i}^t\}$$

and

$$N_2 = E\{\tilde{Y}_{S_j} \tilde{Y}_{S_j}^t\}.$$

Then, as above,

$$N_1 = M[E\{X_{S_i} X_{S_i}^t\} - \mathbf{m}_i \mathbf{m}_i^t] M^t$$

and

$$N_2 = M[E\{X_{S_j} X_{S_j}^t\} - \mathbf{m}_j \mathbf{m}_j^t] M^t$$

where  $\mathbf{m}_k = E\{X_{S_k}\}$  for  $k = i, j$ .

$\Gamma$  is calculated as previously by  $\Gamma = N_1^{-1} N_2$ , and the continuation of the procedure is the same: calculation

of  $P$  and application of  $P^t$  to  $Y$ . This methods experimentally converges to two different sources out of  $N$ . A circular permutation on the indices  $i$  and  $j$  allows to converge to other sources; using the whole combinations of couples  $i, j$ , we retrieve all the sources.

## 5. CALCULATION OF M IN DIFFERENT CASES

### 5.1. Two dimensional case

As shown by equation (8), the values of  $m_1$  and  $m_2$  must be calculated to assure the improvement completed by our algorithm. Unfortunately, this calculation have to be done for each density distribution.

#### 5.1.1. Uniformly distributed sources

If  $X_1$  and  $X_2$  are standardized uniformly distributed ( $\|X_i\| = 1$ ), i.e.  $p_1(\mathbf{x}_1) = p_2(\mathbf{x}_2) \sim U[-\omega; +\omega]$ , calculations can be lead completely. Supposing that the observations have been whitened and that  $|\alpha| > |\beta|$ , the expression of  $m_1$  and  $m_2$  are the following :

$$\begin{cases} m_1 = \frac{1}{2}(1 - \frac{1}{3}(\frac{\beta}{\alpha})^2) \\ m_2 = \frac{1}{3} \frac{\beta}{\alpha} \end{cases} \quad (9)$$

It follows from these expressions that  $|\frac{m_1}{m_2}| > |\frac{\alpha}{\beta}|$  and therefore,  $ind(M^{(2)}) < ind(M^{(1)})$ .

If  $|\alpha| < |\beta|$ , the expressions of  $m_1$  and  $m_2$  are inverted.

At the first step, the separation is rarely reached (except for 2-states signals) but a convergence to the good solution has began.

It can be proved that the iterations lead to a ratio  $|\frac{m_1}{m_2}|$  which is zero or infinite. Then the separation is reached. This result can be extended without any problem to sources taking a limited number of states (with symmetrical probability distributions), if all these states have the same probability to occur.

Practically, during tests done with simulated data, the convergence was always obtained.

#### 5.1.2. Gaussian Distributed Sources

The results (2) to (8) of the previous section stay true.

As  $X_1$  and  $X_2$  are gaussian, i.e.  $p_1(\mathbf{x}_1) = p_2(\mathbf{x}_2) \sim N(0; \sigma^2)$ , in the case of whitened observations, the expressions of  $m_1$  and  $m_2$  are the following :

$$\begin{cases} m_1 = \frac{1}{\sqrt{2\pi}} \sqrt{\frac{\alpha^2}{\alpha^2 + \beta^2}} \\ m_2 = \frac{\beta}{\alpha \sqrt{2\pi}} \sqrt{\frac{\alpha^2}{\alpha^2 + \beta^2}} \end{cases} \quad (10)$$

Obviously  $|\frac{m_1}{m_2}| = |\frac{\alpha}{\beta}|$ . No improvement occurs during the first step of the algorithm.  $M^{(2)}$  is not more discriminating than  $M^{(1)}$  towards the sources. It is not possible to separate the sources.

## 6. SIMULATION RESULTS

### 6.1. Two white sources

A loop of 500 tests where  $M$  is each time a random matrix is performed. To evaluate the performances of the algorithm, we calculate at the end of convergence (triggered by the value of  $|ind(P)|$ ) the performance criterion of  $M$ . Simulations are done upon the following kinds of sources :

- 2 states uniformly distributed sources (2S),
- 4 states uniformly distributed sources (4S),
- 8 states uniformly distributed sources (8S),
- 16 states uniformly distributed sources (16S),
- uniformly distributed sources (unif).

The results are shown in the table below; the columns depict respectively :

- the kind of probability distribution of the sources,
- the mean and the deviation of the final performance criterion of  $M$  noted  $i(M) = ind(M^{(\infty)})$ ,
- the mean of the 2 mean square errors between the estimated sources and the true sources,
- the mean of the number of iterations to converge.

type	$mean(i(M))$	$\sigma_{i(M)}$	mean mse	Ite
2S	$5.13 \times 10^{-5}$	$7.53 \times 10^{-5}$	$3.92 \times 10^{-5}$	1.0
4S	$6.49 \times 10^{-5}$	$7.59 \times 10^{-5}$	$4.66 \times 10^{-5}$	1.8
8S	$9.25 \times 10^{-5}$	$1.09 \times 10^{-4}$	$4.84 \times 10^{-5}$	6.3
16S	$3.56 \times 10^{-4}$	$4.50 \times 10^{-4}$	$8.71 \times 10^{-5}$	6.7
unif	$1.59 \times 10^{-3}$	$4.24 \times 10^{-4}$	$2.06 \times 10^{-4}$	9.0

### 6.2. Robustness of the algorithm

To evaluate how robust is the algorithm towards the number of sources, we use it to separate a growing number of uniformly distributed signals. The iterations are stopped when the mean of the mean square errors between the estimated sources and the true sources is less than  $10^{-4}$  or when the performance index of  $P$  is less than  $10^{-10}$  (that means that an iteration doesn't modify significantly the mixing matrix). The table hereafter shows the number of iterations needed to converge

(100 tests done and the mse obtained are all less than  $10^{-4}$ ):

number of sources	$I$
2	4.8
3	5.4
4	12.8
5	14.4
6	15.7
8	17.9
10	20.6
15	30.1
20	42.3

## 7. CONCLUSION

To our knowledge, the approach presented in this work is innovative. The results obtained on simulated data show that this algorithm is robust towards the number of sources mixed: simulations with more than 20 uniformly distributed sources were successful.

The main advantage of this approach is that it works with sources having the same spectral densities, when classical methods become ineffective.

Modified algorithms of the one presented in this paper have been successfully tested but need still to be totally justified.

## REFERENCES

- [1] L. Tong, V.C. Soon, Y.F. Huang, and R. Liu, *AMUSE: A new blind identification algorithm*, in Proc. 1990 IEEE ISCAS, New Orleans, LA., May 1990
- [2] L. Tong, R. Liu, and V.C. Soon, *Indeterminacy and identifiability of blind identification*, in IEEE Transactions on Circuits and Systems, Vol. 38, No 5, May 1991.
- [3] A. Belouchrani, K. Abed-Meraim, and J.-F. Cardoso, *A blind source separation technique using second order statistics*, in IEEE Transactions on Signal Processing, Vol. 45, No 2, February 1997.
- [4] J.F. Cavassilas, B. Xerri and G. Chabriel, *Séparation autodidacte de sources temporellement corréllées (mélange instantané)*, in GRETSI Symposium, Vol. 1, pp 107-110, Sept. 1997.
- [5] G. Chabriel, B. Xerri and J.F. Cavassilas, *Second Order Blind Identification of Slightly Delayed Mixtures*, in ICA'99 First International Workshop on Independant Component Analysis and Signal Separation, pp 75-79, January 11-15, 1999.

# PERFORMANCE OF CUMULANT BASED INVERSE FILTER CRITERIA FOR BLIND DECONVOLUTION OF MULTI-INPUT MULTI-OUTPUT LINEAR TIME-INVARIANT SYSTEMS

Chong-Yung Chi and Chii-Horng Chen

Department of Electrical Engineering,  
National Tsing Hua University, Hsinchu, Taiwan, R.O.C.  
Tel: 886-3-5731156, Fax: 886-3-5751787, E-mail: cychi@ee.nthu.edu.tw

## ABSTRACT

Tugnait, and Chi and Chen proposed multi-input multi-output inverse filter criteria (MIMO-IFC) using higher-order statistics for blind deconvolution of multi-input multi-output (MIMO) linear time-invariant (LTI) systems. This paper proposes a performance analysis for the MIMO linear equalizer associated with MIMO-IFC for finite SNR, including (P1) perfect phase equalization property, (P2) a relation to MIMO minimum mean square error (MIMO-MMSE) equalizer, and (P3) a connection with the one obtained by Yeh and Yau's MIMO super-exponential algorithm (MIMO-SEA) that usually converges fast but no guarantee of convergence for finite data. Furthermore, based on (P3), a MIMO-IFC based algorithm with performance similar to that of the MIMO-SEA and with guaranteed convergence is proposed. Finally, some simulation results are presented to support the analytic results and the proposed algorithm.

## 1. INTRODUCTION

Blind deconvolution of a multi-input multi-output (MIMO) linear time-invariant system, denoted  $\mathbf{H}[n]$  ( $P \times K$  matrix), is a problem of estimating the vector input  $\mathbf{u}[n] = (u_1[n], \dots, u_K[n])^T$  ( $K$  inputs) with only a set of non-Gaussian vector output measurements  $\mathbf{x}[n] = (x_1[n], \dots, x_P[n])^T$  ( $P$  outputs) as follows [1-3]

$$\mathbf{x}[n] = \sum_{k=-\infty}^{\infty} \mathbf{H}[k] \mathbf{u}[n-k] + \mathbf{w}[n] \quad (1)$$

where  $\mathbf{w}[n]$  ( $P \times 1$  vector) is additive noise. Blind deconvolution of MIMO systems in multiuser detection of wireless communications includes suppression of multiple access interference (MAI) and removal of multiple transmission paths that are crucial to the receiver design of multiuser communications systems.

Let  $\mathbf{v}[n] = (v_1[n], \dots, v_P[n])^T$  denote a linear FIR equalizer of length  $L = L_2 - L_1 + 1$  for which  $\mathbf{v}[n] \neq \mathbf{0}$  for  $n = L_1, L_1 + 1, \dots, L_2$ . Let  $\text{cum}\{y_1, y_2, \dots, y_p\}$  denote the  $p$ th-order cumulant of random variables  $y_1, y_2, \dots, y_p$  and  $\mathcal{F}\{\bullet\}$

denote discrete-time Fourier transform operator. For ease of later use, let us define the following notations

$$\begin{aligned} \text{cum}\{y : p, \dots\} &= \text{cum}\{y_1 = y, \dots, y_p = y, \dots\} \\ C_{p,q}\{y\} &= \text{cum}\{y : p, y^* : q\} \\ \mathbf{v}_j &= (v_j[L_1], \dots, v_j[L_2])^T \\ \boldsymbol{\nu} &= (\mathbf{v}_1^T, \mathbf{v}_2^T, \dots, \mathbf{v}_P^T)^T \\ \mathbf{x}_j[n] &= (x_j[n - L_1], \dots, x_j[n - L_2])^T \\ \mathbf{R}_{i,j} &= E[\mathbf{x}_i^*[n] \mathbf{x}_j^T[n]] \quad (L \times L \text{ matrix}) \\ \tilde{\mathbf{R}} &= \{\mathbf{R}_{i,j}\} \quad (P \times P \text{ block matrix}) \end{aligned}$$

where  $y^*$  denotes the complex conjugate of  $y$ . Then the output  $e[n]$  of the FIR equalizer  $\mathbf{v}[n]$  can be expressed as

$$e[n] = \sum_{j=1}^P v_j[n] * x_j[n] = \sum_{j=1}^P \mathbf{v}_j^T \mathbf{x}_j[n] \quad (2)$$

$$= \sum_{j=1}^K s_j[n] * u_j[n] + w[n] \quad \text{by (1)} \quad (3)$$

where  $w[n]$  is the noise term due to  $\mathbf{w}[n]$  and

$$s_j[n] = \sum_{m=1}^P \sum_{l=L_1}^{L_2} v_m[l] h_{m,j}[n-l] \quad (4)$$

where  $h_{m,j}[n]$  is the  $(m, j)$ th component of  $\mathbf{H}[n]$ . The designed linear equalizer is usually evaluated by the amount of intersymbol interference (ISI) defined as [3, 4]

$$\text{ISI}(e[n]) = \frac{\{\sum_{j,n} |s_j[n]|^2\} - \max_{j,n} \{|s_j[n]|^2\}}{\max_{j,n} \{|s_j[n]|^2\}} \quad (5)$$

Note that  $\text{ISI}(e[n]) = 0$  as  $s_\ell[n] = \alpha \delta[n - \tau]$  and  $s_j[n] = 0$  for  $j \neq \ell$ .

Single-input single-output inverse filter criteria (SISO-IFC) [4-6] using higher-order cumulants have been widely used for blind deconvolution and their performance analyses for finite SNR have been reported by Feng and Chi [5, 6]. In this paper, we propose performance analyses for cumulant based multi-input multi-output inverse filter criteria (MIMO-IFC) [1, 2]. Furthermore, based on the analytic results, a MIMO-IFC based algorithm with performance similar to that of Yeh and Yau's MIMO super-exponential algorithm (MIMO-SEA) [3] and with guaranteed convergence is proposed.

This work was supported by the National Science Council under Grants NSC 88-2218-E007-019 and NSC 89-2213-E007-073.

## 2. REVIEW OF MIMO-IFC AND MIMO-SEA

Assume that we are given a set of measurements  $\mathbf{x}[n]$ ,  $n = 0, 1, \dots, N-1$ , modeled by (1) with the following assumptions:

- (A1)  $u_j[n]$  is zero-mean, independent identically distributed (i.i.d.) non-Gaussian with variance  $\sigma_{u_j}^2$ , and  $(p+q)$ th-order cumulant  $C_{p,q}\{u_j[n]\}$ , and statistically independent of  $u_k[n]$  for all  $k \neq j$ .
- (A2) The MIMO system  $\mathbf{H}[n]$  is exponentially stable.
- (A3) The noise  $\mathbf{w}[n]$  is zero-mean Gaussian and statistically independent of  $\mathbf{u}[n]$ .

Chi and Chen [2] find the optimum  $\nu$  by maximizing the following MIMO-IFC

$$J_{p,q}(\nu) = \frac{|\text{cum}\{e[n] : p, e^*[n] : q\}|}{|\text{cum}\{e[n], e^*[n]\}|^{(p+q)/2}} \quad (6)$$

where  $p$  and  $q$  are nonnegative integers and  $p+q \geq 3$  through using iterative optimization algorithms because all MIMO-IFC  $J_{p,q}$  are a highly nonlinear function of  $\nu$ . Note that the MIMO-IFC given by (6) include Tugnait's MIMO-IFC [1] for  $(p, q) = (2, 1)$  and  $(p, q) = (2, 2)$  as special cases.

The MIMO-SEA proposed by Yeung and Yau [3] iteratively updates  $\nu$  at the  $I$ th iteration by solving the following linear equations

$$\tilde{\mathbf{R}}\nu_I = \frac{1}{\|\tilde{\mathbf{R}}^{-1}\tilde{\mathbf{d}}^{[I-1]}\|} \cdot \tilde{\mathbf{d}}^{[I-1]} \quad (7)$$

where  $\tilde{\mathbf{d}}^{[I-1]} = (\mathbf{d}_1^T, \mathbf{d}_2^T, \dots, \mathbf{d}_P^T)^T$  in which

$$\mathbf{d}_i = \text{cum}\{e^{[I-1]}[n] : r, (e^{[I-1]}[n])^* : s-1, \mathbf{x}_i^*[n]\} \quad (8)$$

in which  $r+s \geq 3$  and  $e^{[I-1]}[n]$  is the equalizer output obtained at the  $(I-1)$ th iteration.

A known fact and two observations regarding MIMO-IFC and MIMO-SEA are as follows:

- (F1) In the absence of noise (i.e.,  $\text{SNR} = \infty$ ), the optimum  $e[n] = \alpha_{\ell} u_{\ell}[n - \tau_{\ell}]$  (perfect equalization) (i.e.,  $\text{ISI}(e[n]) = 0$ ) for both MIMO-IFC and MIMO-SEA as  $L_1 \rightarrow -\infty$  and  $L_2 \rightarrow \infty$  where  $\ell \in \{1, 2, \dots, K\}$  is unknown. For finite SNR and  $L$ ,  $\hat{u}_{\ell}[n] = e[n]$  is an estimate of  $u_{\ell}[n]$  up to a scale factor and a time delay, and  $\hat{h}_{i,\ell}[k]$  can also be estimated as

$$\hat{h}_{i,\ell}[k] = \frac{E[x_i[n+k]\hat{u}_{\ell}^*[n]]}{E[|\hat{u}_{\ell}[n]|^2]}, \quad i = 1, 2, \dots, P \quad (9)$$

- (O1) The computationally efficient MIMO-SEA converges at a super-exponential rate for  $\text{SNR} = \infty$  and sufficiently large  $N$ , but it may diverge for finite SNR and  $N$ .
- (O2) With larger computational load than solving the linear equations given by (7), gradient type iterative MIMO-IFC algorithms (such as Fletcher-Powell algorithm [7]) always spend more iterations (lower convergence speed) than MIMO-SEA.

Estimates  $\hat{u}_1[n], \hat{u}_2[n], \dots, \hat{u}_K[n]$  can be obtained by the MIMO-IFC or MIMO-SEA in a non-sequential order through a multistage successive cancellation (MSC) procedure [1] that includes the following two steps at each stage:

- (S1) Find an input estimate, said  $\hat{u}_{\ell}[n]$  (where  $\ell$  is unknown), and the associated channel estimates  $\hat{h}_{i,\ell}[n]$ ,  $i = 1, 2, \dots, P$  using MIMO-IFC or MIMO-SEA.
- (S2) Update  $x_i[n]$  by  $x_i[n] - \hat{u}_{\ell}[n] * \hat{h}_{i,\ell}[n]$ ,  $i = 1, 2, \dots, P$ .

## 3. PERFORMANCE ANALYSIS FOR MIMO-IFC

Prior to presenting analytical results for the performance of the FIR equalizer  $\mathbf{v}[n]$  associated with MIMO-IFC, let us present the nonblind MIMO minimum mean square error (MIMO-MMSE) equalizer, denoted  $\mathcal{V}_{\text{MMSE}}(\omega)$  ( $K \times P$  matrix), that has some relation to  $\mathbf{v}[n]$ . It can be shown by orthogonality principle [8] that

$$\mathcal{V}_{\text{MMSE}}^T(\omega) = [\mathcal{R}^T(\omega)]^{-1} \cdot \mathcal{H}^*(\omega) \cdot \mathbf{S} \quad (10)$$

where  $\mathcal{R}(\omega) = \mathcal{F}\{\mathbf{R}[k]\} = \mathcal{F}\{E[\mathbf{x}[n]\mathbf{x}^H[n-k]]\}$ ,  $\mathcal{H}(\omega) = \mathcal{F}\{\mathbf{H}[n]\}$  and

$$\mathbf{S} = \text{diag}\{\sigma_{u_1}^2, \dots, \sigma_{u_K}^2\}. \quad (11)$$

Some analytical results regarding the optimum  $\mathbf{v}[n]$  for finite SNR are summarized as follows:

*Property 1.* The optimum overall impulse response  $s_j[n]$  given by (4),  $j = 1, \dots, K$ , are linear phase for finite  $L$ , i.e., their phase responses are given by

$$\arg[S_j(\omega)] = \omega\tau_j + \xi_j, \quad \forall \omega \in [-\pi, \pi] \quad (12)$$

where  $S_j(\omega) = \mathcal{F}\{s_j[n]\}$ ,  $\tau_j$  and  $\xi_j$  are real constants.  $\square$

*Property 2.* The optimum  $\mathbf{V}(\omega) = \mathcal{F}\{\mathbf{v}[n]\}$  for  $L_1 \rightarrow -\infty$  and  $L_2 \rightarrow \infty$  is related to  $\mathcal{V}_{\text{MMSE}}(\omega)$  by

$$\mathbf{V}(\omega) = \mathcal{V}_{\text{MMSE}}^T(\omega) \cdot (\alpha_{p,q} \tilde{\mathbf{S}}_{p,q} \mathbf{D}_{p,q}(\omega) + \alpha_{q,p} \tilde{\mathbf{S}}_{q,p} \mathbf{D}_{q,p}(\omega)) \quad (13)$$

where

$$\alpha_{p,q} = \frac{p \cdot C_{1,1}\{e[n]\}}{(p+q) \cdot C_{q,p}\{e[n]\}}, \quad (14)$$

$$\tilde{\mathbf{S}}_{p,q} = \text{diag}\{C_{q,p}\{u_1[n]\}/\sigma_{u_1}^2, \dots, C_{q,p}\{u_K[n]\}/\sigma_{u_K}^2\} \quad (15)$$

and

$$\mathbf{D}_{p,q}(\omega) = [\mathbf{D}_1(\omega), \dots, \mathbf{D}_K(\omega)]^T \quad (16)$$

in which

$$\mathbf{D}_j(\omega) = \mathcal{F}\{s_j^q[n](s_j^*[n])^{p-1}\}. \quad (17)$$

$\square$

*Property 3.* The optimum  $\mathbf{v}[n]$  and the one obtained by the MIMO-SEA are the same for  $p = q = r = s \geq 2$  and finite  $L$ .  $\square$

Furthermore, based on Property 3 and the observations (O1) and (O2), a fast iterative algorithm is proposed for finding the optimum  $\mathbf{v}[n]$  associated with MIMO-IFC for  $p = q$  as follows:

**Algorithm 1.** Given  $\nu_{I-1}$  and  $e^{[I-1]}[n]$  obtained at the  $(I-1)$ th iteration,  $\nu_I$  at the  $I$ th iteration is obtained by the following two steps.

- (T1) As the MIMO-SEA, obtain  $\nu_I$  by solving (7) with  $r = s = p = q$  and obtain the associated  $e^{[I]}[n]$ .
- (T2) If  $J_{p,p}(\nu_I) > J_{p,p}(\nu_{I-1})$ , go to the next iteration, otherwise update  $\nu_I$  through a gradient type optimization algorithm and obtain the associated  $e^{[I]}[n]$ .

It can be easily shown that

$$\left. \frac{\partial J_{p,p}(\nu)}{\partial \nu} \right|_{\nu=\nu_{I-1}} \propto \frac{1}{C_{p,p}\{e^{[I-1]}[n]\}} \cdot (\tilde{\mathbf{d}}^{[I-1]})^* - \frac{1}{C_{1,1}\{e^{[I-1]}[n]\}} \cdot (\tilde{\mathbf{R}}\nu_{I-1})^* \quad (18)$$

where  $\tilde{\mathbf{d}}^{[I-1]}$  has been obtained in (T1) (see (7)) and  $\tilde{\mathbf{R}}$  is the same at each iteration, indicating simple and straightforward computation for obtaining  $\partial J_{p,p}(\nu)/\partial \nu$  in (T2). Let us conclude this section with the following remark:

- (R1) Algorithm 1 performs as a fast gradient type MIMO-IFC algorithm with convergence speed, computational load, and ISI similar to those of MIMO-SEA (due to the step (T1)) and with guaranteed convergence (due to the step (T2)).

#### 4. SIMULATION RESULTS

A two-input two-output system taken from [1] was considered with the two inputs  $u_1[n]$  and  $u_2[n]$  assumed to be equally probable binary random sequences of  $\{+1, -1\}$ . The synthetic data  $\mathbf{x}[n]$  for  $N = 900$  and  $\text{SNR} = 15$  dB (spatially independent and temporally white Gaussian noise) were processed by the inverse filter  $\mathbf{v}[n]$  of length  $L = 30$  ( $L_1 = 0$  and  $L_2 = 29$ ) associated with MIMO-IFC using the iterative Fletcher-Powell algorithm [7], MIMO-SEA and Algorithm 1, respectively, with  $p = q = r = s = 2$ . The initial condition associated with  $\nu_0$  was  $v_1[n] = v_2[n] = \delta[n - 14]$  for the first stage and  $v_1[n] = \delta[n - 14]$  and  $v_2[n] = 0$  for the second stage of the MSC procedure.

Thirty independent realizations of the optimum  $s_1[n]$  (associated with  $u_1[n]$ ) and the associated thirty ISI versus iteration number obtained at the first stage of the MSC procedure are shown in Figures 1(a) through 1(f) using the three algorithms, respectively. One can see, from Figure 1, that the resultant  $s_1[n]$ 's are linear phase and they are similar for the three algorithms thus verifying Properties 1 and 3, while the convergence speed for the proposed Algorithm 1 is basically the same as that of MIMO-SEA and faster than the MIMO-IFC using Fletcher-Powell algorithm, thus verifying (O2). The corresponding results for  $s_2[n]$  and ISI obtained at the second stage of the MSC procedure are shown in Figures 2(a) through 2(f). These results also support Properties 1 and 3, and (O2), but the MIMO-SEA failed to converge in one realization (see Figure 2(d)) and

the associated  $s_2[n]$  failed to approximate a delta function (see Figure 2(c)) thus verifying (O1). Algorithm 1 outperforms the other two algorithms in the second stage of the MSC procedure because the former converges as fast as the MIMO-SEA in all the thirty realizations (without any divergence) and converges faster than MIMO-IFC using the Fletcher-Powell algorithm.

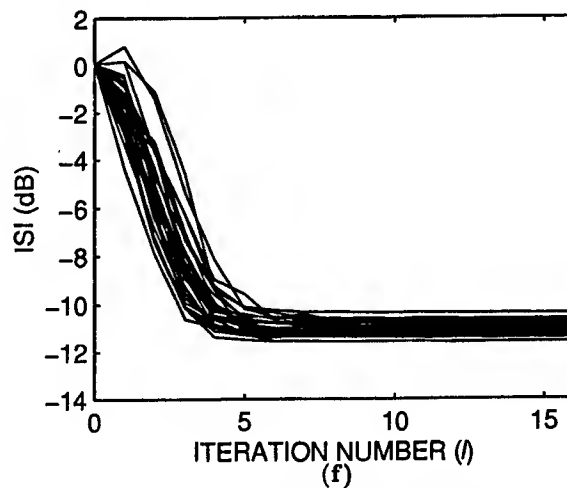
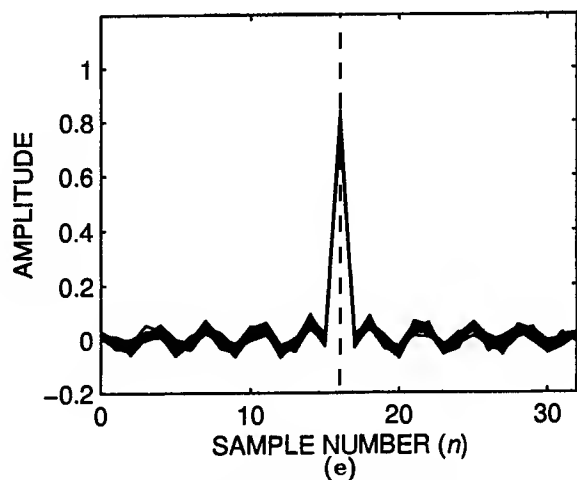
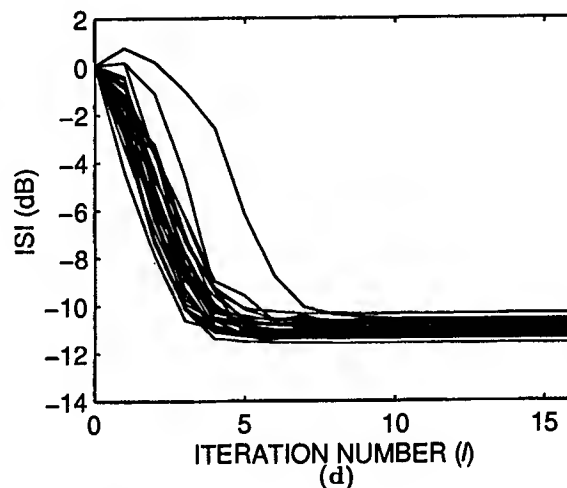
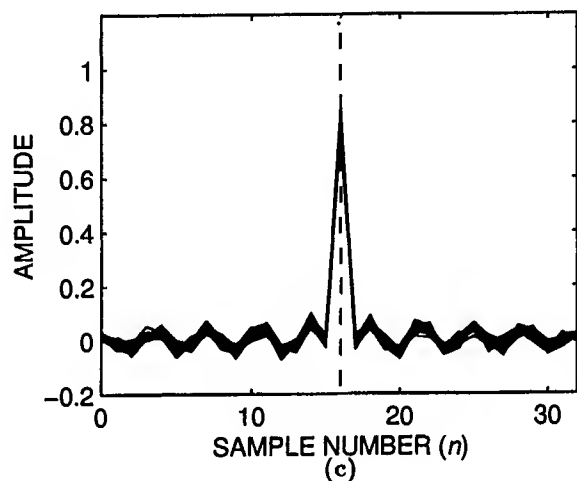
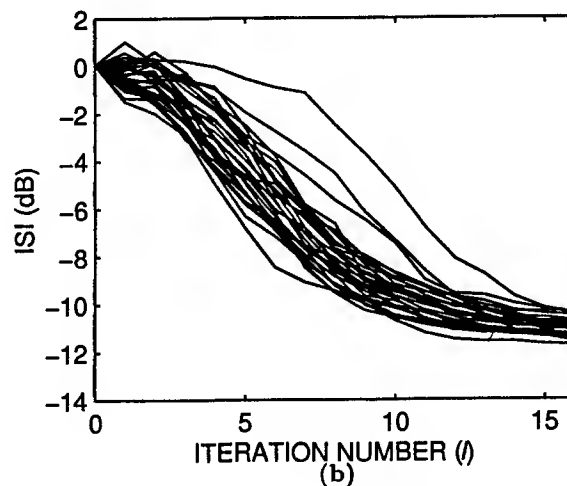
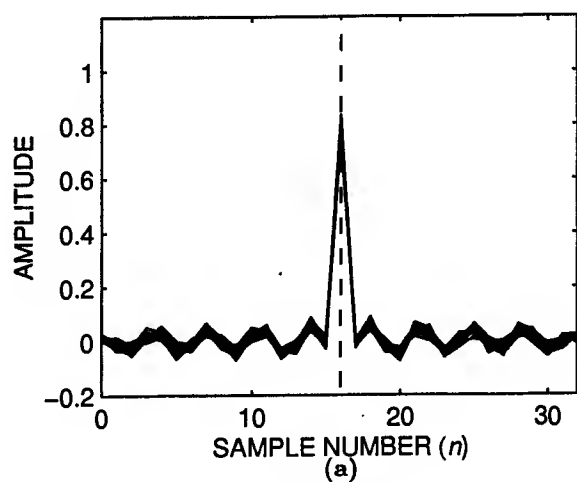
#### 5. CONCLUSIONS

We have presented a performance analysis for the MIMO linear equalizer  $\mathbf{v}[n]$  associated with Chi and Chen's MIMO-IFC for finite SNR, including perfect phase equalization, a relation to the nonblind MIMO-MMSE equalizer, and equivalence to the one associated with MIMO-SEA for  $p = q = r = s$ , as presented in Properties 1, 2 and 3 respectively. Based on Property 3, a MIMO-IFC based algorithm, Algorithm 1, was presented that performs as the MIMO-SEA (in terms of ISI, computational load and convergence speed) with guaranteed convergence (see (R1)) while the latter may not converge for finite SNR and data (see (O1)). Some simulation results were also presented that support the proposed analytical results and Algorithm 1. The application of MIMO-IFC to multiuser detection of CDMA systems using Algorithm 1 is under study.

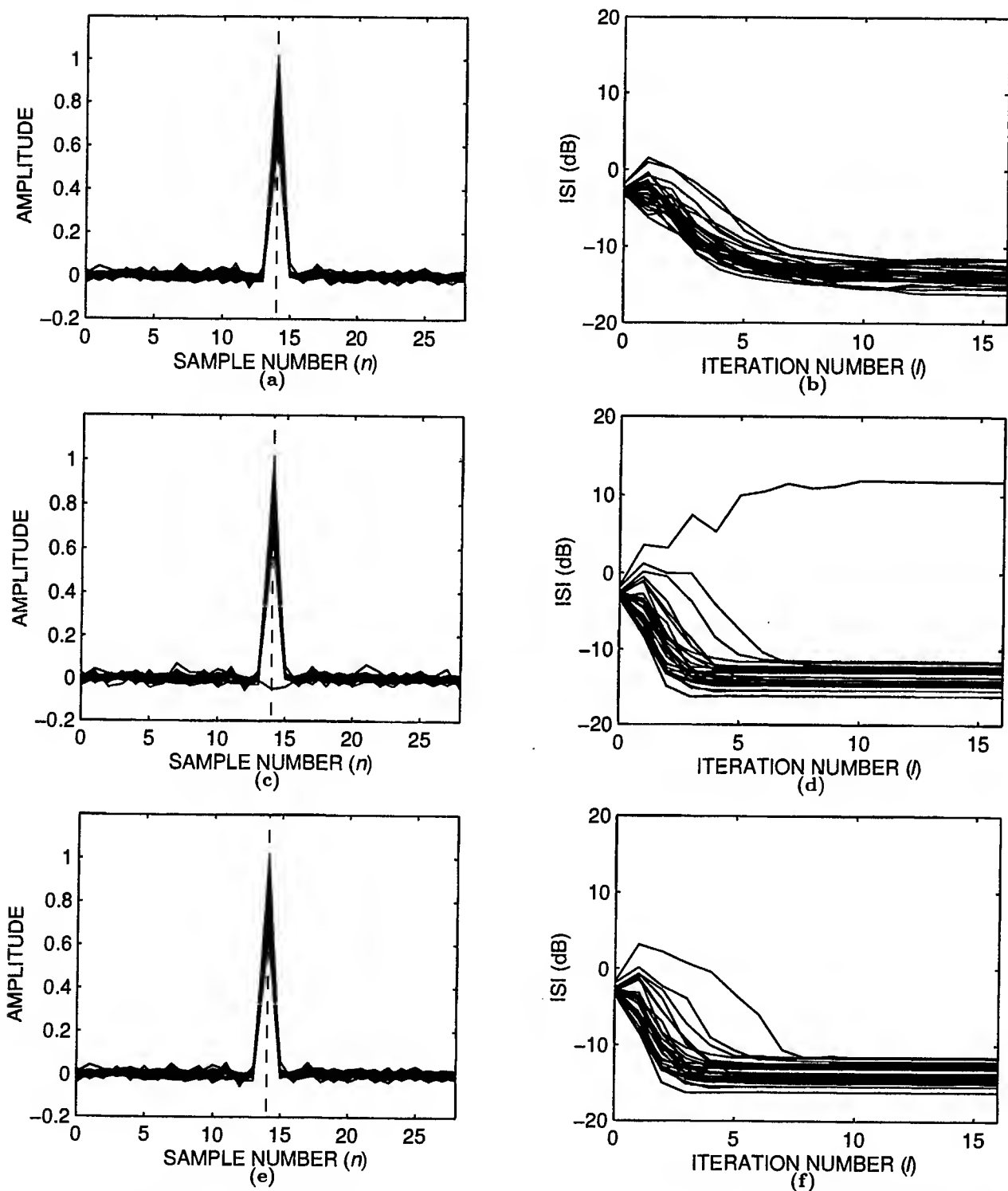
#### 6. REFERENCES

- [1] J. K. Tugnait, "Identification and deconvolution of multichannel linear nonGaussian processes using higher-order statistics and inverse filter criteria," *IEEE Trans. Signal Processing*, vol. 45, no. 3, pp. 658-672, March 1997.
- [2] C.-Y. Chi and C.-H. Chen, "Blind MAI and ISI suppression and channel estimation for DS/CDMA systems using HOS based inverse filter criteria," submitted to *IEEE Trans. Signal Processing*.
- [3] K. L. Yeung and S. F. Yau, "A cumulant-based super-exponential algorithm for blind deconvolution of multi-input multi-output systems," *Signal Processing*, vol. 67, no. 2, pp. 141-162, 1998.
- [4] O. Shalvi and E. Weinstein, *Universal Methods for Blind Deconvolution*, A chapter in S. Haykin, ed., *Blind Deconvolution*, Prentice-Hall, Englewood Cliffs, New Jersey, 1994.
- [5] C.-C. Feng and C.-Y. Chi, "Performance of cumulant based inverse filters for blind deconvolution," *IEEE Trans. Signal Processing*, vol. 47, no. 7, pp. 1922-1935, July 1999.
- [6] C.-C. Feng and C.-Y. Chi, "Performance of Shalvi and Weinstein's deconvolution criteria for channels with/without zeros on the unit circle," *IEEE Trans. Signal Processing*, vol. 48, no. 2, pp. 571-575, Feb. 2000.
- [7] D. M. Burley, *Studies in Optimization*, Falsted Press, a Division of John Wiley and Sons, Inc., New York, 1974.
- [8] C. W. Therrien, *Discrete-Time Random Signals and Statistical Signal Processing*, Englewood Cliffs, Prentice-Hall, 1992.





**Fig. 1.** Thirty simulation results of  $s_1[n]$  and ISI versus iteration number  $I$  at the first stage of the MSC procedure. (a)  $s_1[n]$  and (b) ISI associated with MIMO-IFC for  $p = q = 2$  using Fletcher-Powell Algorithm, (c)  $s_1[n]$  and (d) ISI associated with MIMO-SEA for  $r = s = 2$ , and (e)  $s_1[n]$  and (f) ISI associated with Algorithm 1 for  $p = q = r = s = 2$ .



**Fig. 2.** Thirty simulation results of  $s_2[n]$  and ISI versus iteration number  $I$  at the second stage of the MSC procedure. (a)  $s_2[n]$  and (b) ISI associated with MIMO-IFC for  $p = q = 2$  using Fletcher-Powell Algorithm, (c)  $s_2[n]$  and (d) ISI associated with MIMO-SEA for  $r = s = 2$ , and (e)  $s_2[n]$  and (f) ISI associated with Algorithm 1 for  $p = q = r = s = 2$ .

# SEPARATION OF NON STATIONARY SOURCES; ACHIEVABLE PERFORMANCE

Jean-François Cardoso

C.N.R.S. / E.N.S.T., département TSI  
46, rue Barrault / 75634 Paris Cedex 13 / France  
<http://tsi.enst.fr/cardoso/~stuff.html>

## ABSTRACT

We consider the blind separation of an instantaneous mixture of non stationary source signals, possibly normally distributed. The asymptotic Cramér-Rao bound is exhibited in the case of known source distributions: it reveals how non stationarity and non Gaussianity jointly governs the achievable performance via an index of non stationarity and an index of non Gaussianity.

## 1. INTRODUCTION

The problem of blind separation of instantaneous mixtures is most often addressed by exploiting the possible non Gaussianity of the sources. Actually, this is the only possible route when the source signals are independently and identically distributed (i.i.d.) [3]. As a corollary, i.i.d. sources can be separated only when they are not normally distributed. When the first 'i' of 'i.i.d.' is not valid, *i.e.* when the source signals are correlated in time, another route is to exploit these correlations; identifiability is granted provided the source signals have different spectra (see *e.g.* [6, 1] for more elaborate statements and some algorithms) even when the signals are normally distributed.

In this paper, we consider the case when the second 'i' of 'i.i.d.' is invalid, that is, we set out to achieve signal separation by exploiting the variation in distribution of the source signals. Essentially, we consider the problem of separating *non stationary* signals.

The first section describes the model of interest in its simplest possible form and gives the estimating equations of the maximum likelihood estimator. The second section gives the results of an asymptotic analysis for the asymptotically achievable accuracy of separation in the simple case when the source distributions are known in advance. This case never occurs in practice but the results of the analysis provide us with an upper bound to the performance; it also shows what is the measure of non stationarity which governs blind separability. A final section of the manuscript outlines the implementation of novel source separation algorithms based on the Gaussian non stationary model.

## 2. LIKELIHOOD

We consider a source separation model which is as simple as possible but still lets the non stationary features appear as clearly as possible: an instantaneous mixture where  $T$  samples of a random  $n$ -vector  $\mathbf{x}(t)$  are represented as the mixture by an invertible  $n \times n$  matrix  $A$  of  $T$  samples of a 'source vector'  $\mathbf{s}(t)$ :

$$\mathbf{x}(t) = A\mathbf{s}(t), \quad 1 \leq t \leq T. \quad (1)$$

The model for the joint distribution of  $\mathbf{x}(1), \dots, \mathbf{x}(T)$  is specified as soon as we specify the joint distribution of  $\mathbf{x}(1), \dots, \mathbf{x}(T)$ . Before describing our approach, we first show how blind identifiability can stem from non stationarity in a simple case.

**A simple case.** We first give a very simple example to show that blind separation is possible even for Gaussian sources thanks to non stationarity. Let us assume for instance that the data points are observed during two different regimes: during a first period, the covariance matrix of the source vector is a diagonal matrix  $\Delta_1$  then, during a second period it is a *different* diagonal covariance matrix  $\Delta_2$ . If these two periods are known, one can compute the sample covariance matrix of the observations over each of them, yielding estimates of  $R_1 = A\Delta_1A^\dagger$  and of  $R_2 = A\Delta_2A^\dagger$ . This particular structure determines almost completely matrix  $A$  since the columns of  $A$  are the eigenvalues of  $R_1R_2^{-1}$ . These are generally unique (up to the usual indeterminations of permutation and scale). Denoting  $B = A^{-1}$ , we also remark that  $BR_iB^\dagger = \Delta_i$  for  $i = 1, 2$ : matrix  $B$  jointly diagonalizes the two covariance matrices. This line of reasoning was followed in [9] and [8]. See section 4 to see how, under more general assumptions, the Gaussian log-likelihood turns out to be a joint diagonalization criterion of covariance matrices.

**Maximum likelihood.** We shall now consider the maximum likelihood solution for model (1) when the source distributions are known. The model for the sequence of source signals is as follows. The sequence  $\{\mathbf{s}(t)\}$  is *not* modeled as i.i.d. (which implies stationarity) but as 'Temporally Independently Distributed', abbreviated in 't.i.d.' in the following, and meaning that  $\mathbf{s}(t)$  is distributed independently from  $\mathbf{s}(t')$  for  $t \neq t'$ . In addition, we maintain the usual assumption that the components of the source vector  $\mathbf{s}(t)$  are mutually independent for each  $t$ . Therefore, denoting by  $r_{ti}$  the density of the  $i$ -th component of  $\mathbf{s}(t)$ , the density of a sample  $\mathbf{s}(1), \dots, \mathbf{s}(T)$  is the product:

$$p(\mathbf{s}(1), \dots, \mathbf{s}(T)) = \prod_{t=1}^T \prod_{i=1}^n r_{ti}(s_i(t)) \quad (2)$$

In this 't.i.d. model', the relative gradient [2] of the log-likelihood of  $T$  data points is easily found to be

$$-\nabla \log p(\mathbf{x}(1), \dots, \mathbf{x}(T)|A) = \sum_{t=1}^T \left( \phi_t(\mathbf{y}(t))\mathbf{y}(t)^\dagger - I \right) \quad (3)$$

where  $I$  denotes the identity matrix, where

$$\mathbf{y}(t) = A^{-1}\mathbf{x}(t) \quad (4)$$

and where  $\phi_t(\cdot)$  is the vector-to-vector mapping:

$$[\phi_t(\mathbf{y})]_i = \phi_{ti}(\mathbf{y}) = -\frac{r'_{ti}(y_i)}{r_{ti}(y_i)}. \quad (5)$$

This is obtained in a straightforward generalization of the i.i.d. case.

**Specializing to a Gaussian non stationary model.** In the following, special attention is paid to the Gaussian case: at time  $t$ , the  $i$ -th source is drawn according to a zero-mean Gaussian distribution with variance  $\sigma_{ti}^2$ . This is  $\mathbf{s}(t) \sim \mathcal{N}(0, \Delta_t)$  where the diagonal covariance matrix at time  $t$  is

$$\Delta_t = \text{diag}(\sigma_{t1}^2, \dots, \sigma_{tn}^2). \quad (6)$$

In this case, the score functions  $\phi_{ti}$  are the linear functions:

$$\phi_{ti}^g(\mathbf{y}) = \frac{y_i}{\sigma_{ti}^2} \quad (7)$$

or, matrix-wise:

$$\phi_t^g(\mathbf{y}) = \Delta_t^{-1} \mathbf{y}. \quad (8)$$

Combining eq. (3) and eq. (7), the stationary points of the likelihood are—in the Gaussian case—the solutions of the estimating equations:

$$\frac{1}{T} \sum_{t=1}^T \frac{y_i(t)y_j(t)}{\sigma_{ti}^2} - \delta_{ij} = 0 \quad 1 \leq i, j \leq n. \quad (9)$$

Of course, if the variances are assumed to be constant, ( $\sigma_{ti} = \sigma_i$ ), this set of equations becomes redundant: the  $(i, j)$ -th term provides us with the same condition as the  $(j, i)$ -th term and the model is not identifiable. In contrast, in the non stationary case, these two conditions are *a priori* distinct and the set (9) of equations yields a number of constraints equal to the number of unknown parameters in  $A$ . A similar set of estimations has been derived in [4] without reference to the maximum likelihood principle.

### 3. ACHIEVABLE PERFORMANCE IN SEPARATING NON GAUSSIAN NON STATIONARY SOURCES

In this section, we compute the Fischer information matrix in the t.i.d. case when the source distributions are known at each time instant. By this device, we obtain an expression for the asymptotic Cramér-Rao bound which can be simply and directly related to the achievable separation and to the non stationarity and non Gaussianity of the sources.

#### 3.1. Non stationary averages.

The asymptotic derivations developed by Pham [7] for the stationary case can be generalized to the non stationary case. However, some statistical moments must receive a more general definition, adapted to the non stationary case: the mathematical expectation operator  $E$  must be replaced by a limit of expected values. This will be denoted by the operator  $\bar{E}$  defined as follows. If  $\{X_t\}$  is a sequence of random variables with the distribution of  $X_t$  depending on  $t$ , we write

$$\bar{E}\{X\} \stackrel{\text{def}}{=} \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T EX_t, \quad (10)$$

assuming that such a limit actually exists (if it does not, it becomes difficult to carry any asymptotic analysis...). In particular, the average power of the  $i$ -th signal is:

$$\bar{E}\{s_i^2\} = \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=1}^T Es_i(t)^2. \quad (11)$$

#### 3.2. Rejection rates.

The accuracy of an estimate  $\hat{A}$  of  $A$  in terms of source separation can be measured by the associated 'rejection rates'. If the source vector is estimated as  $\hat{\mathbf{s}} = \hat{A}^{-1}\mathbf{x}$ , then  $\hat{s}_i = \sum_j [\hat{A}^{-1}A]_{ij} s_j$  so that the average power of the  $j$ -th source in the estimate of the  $i$ -th source is

$$[\hat{A}^{-1}A]_{ij}^2 \bar{E}\{s_j^2\} \quad (12)$$

while the average power of the  $i$ -th source itself in the same estimate is

$$[\hat{A}^{-1}A]_{ii}^2 \bar{E}\{s_i^2\}. \quad (13)$$

For a regular estimator, the asymptotic variance of the estimate  $\hat{A}$  is expected to decrease as  $T^{-1}$  so that  $[\hat{A}^{-1}A]_{ij}^2$  is of order  $1/T$  for  $i \neq j$  while  $[\hat{A}^{-1}A]_{ii}$  converges to the constant 1. Therefore, a significant characterization of the accuracy of a given estimator is obtain by evaluating the *asymptotic rejection rates*:

$$\rho_{ij} \stackrel{\text{def}}{=} \lim_{T \rightarrow \infty} TE \left( [\hat{A}^{-1}A]_{ij}^2 \right) \frac{\bar{E}\{s_j^2\}}{\bar{E}\{s_i^2\}} \quad (14)$$

which are nothing but properly scaled interference-to-signal ratios.

#### 3.3. The Fisher information matrix.

The computation of Fisher information matrices (FIMs) in source separation is facilitated by resorting to the relative gradient. It is equivalent to a local re-parameterization in term of a relative (or multiplicative) variation: in order to compute the FIM at point  $A$ , any matrix in the neighborhood of  $A$  is expressed as  $A(I + \mathcal{E})$  where matrix  $\mathcal{E}$  is the 'local parameter' or 'relative parameter'. The 'relative score' is the derivative of the log-likelihood with respect to  $\mathcal{E}$  evaluated at  $\mathcal{E} = 0$ , that is the matrix with entries:

$$g_{ij} = \frac{\partial}{\partial \mathcal{E}_{ij}} \log(p(\mathbf{x}(1), \dots, \mathbf{x}(T) | A(I + \mathcal{E}))) \Big|_{\mathcal{E}=0} \quad (15)$$

Similar to eq. (3), the relative score is found to be

$$g_{ij} = - \sum_{t=1}^T \{ \phi_{ti}(s_i(t)) s_j(t) - \delta_{ij} \} \quad (16)$$

In the t.i.d. model, thanks to the independence assumptions, it is not difficult to evaluate the 'relative FIM', i.e. the covariance matrix of the relative scores. One finds that for  $i \neq j$ ,  $g_{ij}$  is only correlated to  $g_{ji}$  (and to itself!) and that  $g_{ii}$  is uncorrelated to  $g_{kl}$  unless  $i = k = l$ . Therefore the relative FIM is an  $n^2 \times n^2$  matrix which is block diagonal: there are  $n(n-1)/2$  blocks of size  $2 \times 2$  for each pair  $1 \leq i < j \leq n$  of sources which are the covariance matrices of  $[g_{ij}, g_{ji}]^T$  and  $n$  blocks of size  $1 \times 1$  for each source, equal to  $Eg_{ii}^2$ ,  $1 \leq i \leq n$ . The  $2 \times 2$  blocks are the most interesting. Using independence and the zero-mean assumption, one readily finds that for  $i \neq j$ ,

$$Eg_{ij}^2 = \sum_t E\phi_{ti}^2(s_i(t)) Es_j^2(t) \quad (17)$$

$$Eg_{ij}g_{ji} = T \quad (18)$$

where we have used that  $E\phi_{ti}(s_i(t))s_i(t) = 1$ . This completes the computation of the Fisher information matrix but does not provide many insights. More interesting expressions are obtained by relating the FIM to the rejection rates and introducing the additional assumption of 'independent non-stationarities' described at next section.

### 3.4. Independent non stationarities.

In our model, the distributions of the sources at each time instant are fixed. We cannot expect meaningful result without some kind of assumption expressing that the *distributions* of the sources are 'independent' themselves. We will consider the case where

$$\bar{E}\{\phi_i^2(s_i)s_j^2\} = \bar{E}\{\phi_i^2(s_i)\} \bar{E}\{s_j^2\}. \quad (19)$$

This condition must be understood as some 'independence of the non stationarities' since it expresses that the sequence  $\{E\phi_i^2(s_i(t))\}$  is 'uncorrelated' with the sequence  $\{Es_j^2(t)\}$  of the variances of the  $j$ -th source. The word 'uncorrelated' is quoted in the previous sentence because, in the model under consideration, these sequences are not random sequences: one should rather talk of a limiting empirical decorrelation between them. Of course, since these two sequences refer to two different sources, this 'empirical decorrelation' condition (19) is expected to hold in many practical situations as soon as two source signals originate from physically independent processes.

Under this assumption (19), the expression (17) for  $Eg_{ij}^2$  produces the limiting form

$$\lim_{T \rightarrow \infty} T^{-1} Eg_{ij}^2 = (R_i + 1) \frac{\bar{E}\{s_j^2\}}{\bar{E}\{s_i^2\}} \quad (20)$$

where we have defined the scalars  $R_1, \dots, R_n$  as

$$R_i \stackrel{\text{def}}{=} \bar{E}\{\phi_i^2(s_i)\} \bar{E}\{s_i^2\} - 1. \quad (21)$$

Therefore, the limit for the  $2 \times 2$  sub-block of the FIM corresponding to the  $(i, j)$ -th pair of sources is given by

$$\lim_T T^{-1} \text{Cov} \left( \begin{bmatrix} g_{ij} \\ g_{ji} \end{bmatrix} \right) = \begin{bmatrix} (R_i + 1) \frac{\bar{E}\{s_j^2\}}{\bar{E}\{s_i^2\}} & 1 \\ 1 & (R_j + 1) \frac{\bar{E}\{s_i^2\}}{\bar{E}\{s_j^2\}} \end{bmatrix} \quad (22)$$

### 3.5. The Cramér-Rao bound

We have obtained at eq. (22) an asymptotic expression for a  $2 \times 2$  sub-block of the FIM for the relative parameter  $\mathcal{E}$ . Because the FIM is block-diagonal, it suffices to invert these sub-blocks to obtain the large sample Cramér-Rao bound (CRB) for this parameter. In particular, the upper left entry of the inverse of the right hand side of (22) is

$$\frac{R_j + 1}{(R_i + 1)(R_j + 1) - 1} \cdot \frac{\bar{E}\{s_i^2\}}{\bar{E}\{s_j^2\}} \quad (23)$$

and equals (within a factor  $T$ ) the lowest variance  $E\mathcal{E}_{ij}^2$  asymptotically achievable by an unbiased estimate of the relative parameter  $\mathcal{E}$ .

Note that if an estimate  $\hat{A}$  is parameterized as  $A(I + \mathcal{E})$  then, at first order,  $\hat{A}^{-1}A = I - \mathcal{E}$ . Therefore the CRB on the relative parameter  $\mathcal{E}$  is directly related to the rejection rates defined at

(14). In particular, using (23), we find that the best asymptotically achievable rejection rates are

$$\rho_{ij} = \frac{R_j + 1}{R_i R_j + R_i + R_j}. \quad (24)$$

### 3.6. Non stationarity and non Gaussianity

It is important to note that the bound (24) is obtained without assuming neither that the source signals are Gaussian nor that they are stationary. Therefore, expression (24) gives a unifying answer to the issue of finding how the non stationarity *and* the non Gaussianity jointly govern the achievable performance of source separation.

Since  $R_i \geq 0$  (see eq. (30) below), it is more instructive to rewrite (24) as a signal-to-interference ratio:

$$\frac{1}{\rho_{ij}} = R_i + \frac{R_j}{R_j + 1} \quad (25)$$

because this last expression makes it clear that good performance depends on having the highest possible values for  $R_i$  and  $R_j$ . Conversely, performance is at its worst (and blind separation in the t.i.d. case becomes impossible) for a given pair  $(i, j)$  of sources when  $R_i = R_j = 0$ . It is thus important to understand the meaning of the  $R_i$  moments.

**Non Gaussianity index.** Consider the following moment of  $s_i(t)$ :

$$\gamma_{ti} = E\phi_{ti}^2(s_i(t)) Es_i^2(t) - 1. \quad (26)$$

The scalar  $\gamma_{ti}$  is non negative:  $\gamma_{ti} \geq 0$  with equality only when  $s_i(t)$  has a Gaussian distribution. This is easily seen using the Cauchy-Schwartz inequality and the fact that  $E\phi_{ti}(s_i(t))s_i(t) = 1$ . Thus  $\gamma_{ti}$  is a measure of the *non Gaussianity* of the variable  $s_i(t)$ . We define for the  $i$ -th source an 'average' non Gaussianity index:

$$\alpha_i = \frac{\bar{E}\{\gamma_{ti} \sigma_{ti}^{-2}\}}{\bar{E}\{\sigma_{ti}^{-2}\}} \quad (27)$$

as the average over time of the non Gaussianity  $\gamma_{ti}$  of  $s_i(t)$  weighted by the reciprocal variance  $\sigma_{ti}^{-2}$ .

**Non stationarity index.** We defined a non stationarity index for the  $i$ -th source by

$$\beta_i = (\bar{E}\{\sigma_{ti}^2\}) (\bar{E}\{\sigma_{ti}^{-2}\}) - 1. \quad (28)$$

An alternative expression for this index is

$$\beta_i = \bar{E} \left( \frac{\sigma_{ti}}{\sqrt{E\sigma_{ti}^2}} - \frac{\sqrt{E\sigma_{ti}^2}}{\sigma_{ti}} \right)^2 \quad (29)$$

which shows that  $\beta_i \geq 0$  with the equality case being  $\sigma_{ti}^2 = \bar{\sigma}_{ti}^2$ . Thus,  $\beta_i$  does work as a measure of non stationarity or, more accurately, as a measure of second-order stationarity.

**Non stationarity and non Gaussianity.** The moment  $R_i$  which represents the combined effects of non stationarity and non Gaussianity can be rewritten, at the cost of a few manipulations, as a function of  $\alpha_i$  and  $\beta_i$ . We find

$$R_i = \alpha_i + \beta_i + \alpha_i \beta_i \quad (30)$$

In the case of *stationary non Gaussian* sources, we have  $\gamma_{ti} = 0$  so that  $\beta_i = 0$  and  $R_i$  reduces to  $R_i = \alpha_i$  while  $\alpha_i$  itself reduces

to  $\alpha_i = E\phi_i^2(s_i)Es_i^2$ ; the results of Pham [7] for the stationary case are recovered with the usual measure of non Gaussianity.

In the case of *non stationary Gaussian* sources, the score function is given by eq. (7). Then  $\gamma_{ti} = 0$  and thus  $\alpha_i = 0$  so that  $R_i$  reduces to  $R_i = \beta_i$  showing that it is the particular way in which  $\beta_i$  measures the deviation of the variance sequence from being constant which quantifies the potential of non stationarity for blind separation.

Also note that when the sources are weakly non Gaussian ( $\alpha_i \ll 1$ ) and weakly non stationary ( $\beta_i \ll 1$ ), then  $R_i \approx \alpha_i + \beta_i$ , i.e. the benefits of non Gaussianity and non stationarity just *add up*. On the opposite side, for sources which are strongly non stationary and non Gaussian, we have  $R_i \approx \alpha_i\beta_i$ , i.e. the benefits of non Gaussianity and non stationarity *multiply* each other to a large value of  $R_i$ .

#### 4. ALGORITHMS

We briefly outline algorithms for the separation of Gaussian non stationary sources derived from the maximum likelihood principle. More details about the algorithms can be found in [5]

##### 4.1. Gradient and Newton-like algorithms

**Relative gradient algorithm.** The relative gradient algorithm for maximizing the likelihood of a separating matrix  $B = A^{-1}$  in the non stationary case is a direct generalization of the i.i.d. case:

1. Initialize :  $B = I$  (for instance).
2. Compute the relative gradient  $G$  of the log-likelihood:

$$G(B) = \frac{1}{T} \sum_{t=1}^T \phi_t(y(t))y(t)^\dagger - I \quad (31)$$

3. If matrix  $G(B)$  is small enough, stop; otherwise, update the separating matrix  $B$ :

$$B \leftarrow (I - \mu G(B))B \quad (32)$$

and go to step 2.

In practical cases, one cannot expect to know in advance the time-varying distributions of the sources: a ML algorithm cannot be directly implemented as summarized by eqs. (31–32). In the stationary case, an option is to use prior estimates for the non linear score functions  $\phi_i$  or to estimate them from the data using Pham's method [7]. In the non stationary case, there is another option because one does not need to exploit the non Gaussianity: it is sufficient to rely on the non stationarity. Therefore, one may use the Gaussian score functions of eq. (7). These score functions depend on a single parameter: the variance of the given source at the given time instant. Of course, this is still as many parameters as data points so there is no way the instantaneous variances can be estimated without additional assumptions. This suggests a class of algorithms where each iteration of the algorithm (31–32) is intertwined with an estimation of the source variances from the current estimates  $y(t) = Bx(t)$  of the sources.

There are as many algorithms as models and estimation techniques for the variances. The most natural and generic approach is to assume some smoothness in the temporal evolution of the variances of each source. The simplest idea then is to estimate the variances  $\sigma_{ii}^2$  by low-pass filtering the squared outputs  $y_{ii}^2$  of the separating matrix.

**A Newton-like algorithm.** Even though the relative gradient algorithm outlined above performs reasonably well, there is room for simple improvements by developing an approximate on line Newton-like technique. The starting point is an exponentially weighted relative gradient:

$$\bar{G}(t) = \sum_{\tau \leq t} \mu(1 - \mu)^{t-\tau} \left\{ \phi_\tau(y(\tau))y(\tau)^\dagger - I \right\}. \quad (33)$$

Assume that  $\bar{G}(t-1) = 0$  for some  $B(t-1)$  and look for a relative update, that is,  $B(t) = (I - \mu H(t))B(t-1)$ . A first order (in  $\mu$ ) expansion of the equation  $\bar{G}(t) = 0$  and some simplifying assumptions allow to derive a scale-invariant adaptive algorithm which converges much faster than a 'regular' relative gradient technique at very small additional cost.

##### 4.2. Joint diagonalization algorithms

Here, we consider a different model of non stationarity: rather than assuming a smooth temporal variation of the variance profiles, we assume that data set can be divided in  $L$  blocks with the variance of each source being constant over each block. It must be stressed that we do not really need this model to hold to get consistent estimates: it suffices that it captures enough of the non stationarities.

Under this 'piecewise stationary model', the normalized log likelihood takes the form

$$-\frac{1}{T} \nabla \log p(x(1), \dots, x(T)|A) \stackrel{c}{=} \sum_{l=1}^L w_l \text{off}(B \hat{R}_l B^\dagger) \quad (34)$$

where  $\stackrel{c}{=}$  means 'equal up to a constant term', where  $B = A^{-1}$ , where  $\hat{R}_l$  denotes the sample covariance matrix of the observations estimated over the  $l$ -th block, where  $w_l$  is the proportion of data points belonging to the  $l$ -th block and where  $\text{off}(R)$  is a measure of diagonality of a symmetric positive matrix  $R$  defined as

$$\text{off}(R) = \log \det R - \log \det \text{diag} R \quad (35)$$

with  $\text{diag} R$  denoting the diagonal matrix with the same diagonal as  $R$ .

Thus, in the piecewise stationary model, the ML principle boils down to estimating  $A$  as the matrix whose inverse jointly diagonalizes the sample covariances over each block (only an *approximate* joint diagonalization in the weighted sense of eq. (34) is possible, of course).

Based on this principle, one could implement a two-step procedure: a first whitening step which turns the mixing matrix  $A$  into an (approximately) orthogonal matrix followed by a second step of joint approximate diagonalization of the covariance matrices of the whitened data by an orthogonal matrix. This technique would be the non stationary counterpart of [1] (which uses a *spectral* contrast as opposed to the current non-stationary contrast).

However it is possible to implement a better solution because it exists an efficient algorithm for the joint diagonalization of several symmetric positive matrices. This algorithm minimizes exactly the objective (34) and does so over all invertible matrices. Thus it does not require pre-whitening and computes exactly the ML estimate of the piecewise stationary model. More details are in [5].

## 5. CONCLUSION

We have investigated the achievable blind source separation performance in a simple model of non stationary sources. In this model, blind separation is made possible by non stationarity and/or non Gaussianity. The performance is summarized by the rejection rates  $\rho_{ij}$  which under a natural simplifying assumption depend on the moments  $R_i$ . Better performance is obtained for larger  $R_i$ . These moments in turn are simple increasing functions of a non Gaussianity index  $\alpha_i$  and a non stationarity index  $\beta_i$ . Several algorithms for the separation of Gaussian non stationary sources have been outlined; the analysis of their performance will be the subject of further research.

## 6. REFERENCES

- [1] Adel Belouchrani, Karim Abed Meraim, Jean-François Cardoso, and Éric Moulines. A blind source separation technique based on second order statistics. *IEEE Trans. on Sig. Proc.*, 45(2):434–44, February 1997.
- [2] Jean-François Cardoso and Beate Laheld. Equivariant adaptive source separation. *IEEE Trans. on Sig. Proc.*, 44(12):3017–3030, December 1996.
- [3] P. Comon. Independent component analysis, a new concept ? *Signal Processing, Elsevier*, 36(3):287–314, April 1994. Special issue on Higher-Order Statistics.
- [4] K. Matsuoka, M. Ohya, and M. Kawamoto. A neural net for blind separation of nonstationary signals. *Neural networks*, 8(3):411–419, 1995.
- [5] Dinh-Tuan Pham and Jean-François Cardoso. Blind separation of instantaneous mixtures of non stationary sources. In *Proc. ICA 2000, Helsinki, Finland, 2000*. To appear.
- [6] Dinh-Tuan Pham and Philippe Garat. Séparation aveugle de sources temporellement corrélées. In *Proc. GRETSI*, pages 317–320, 1993.
- [7] Dinh-Tuan Pham and Philippe Garat. Blind separation of mixture of independent sources through a quasi-maximum likelihood approach. *IEEE Tr. SP*, 45(7):1712–1725, July 1997.
- [8] Antoine Souloumiac. Blind source detection and separation using second order nonstationarity. In *Proc. ICASSP*, pages 1912–1915, 1995.
- [9] Michail K. Tsatsanis and Changyeul Kweon. Source separation using second order statistics: Identifiability conditions and algorithms. In *Proc. 32nd Asilomar Conf. on Signals, Systems, and Computers*, pages 1574–1578. IEEE, November 1998.

# MODIFIED BSS ALGORITHMS INCLUDING PRIOR STATISTICAL INFORMATION ABOUT MIXING MATRIX

*Jorge Igual, Luis Vergara*

Departamento Comunicaciones  
Universidad Politécnica Valencia  
Camino de Vera, S/N, Valencia 46008, Spain  
e-mail: [jigual@dcom.upv.es](mailto:jigual@dcom.upv.es); [lvergara@dcom.upv.es](mailto:lvergara@dcom.upv.es)

## ABSTRACT

In Blind Source Separation (BSS) no prior knowledge is considered. However, due to inherent indeterminations of the problem, some arbitrary normalizing conditions are imposed on the sources or on the recovering matrix. We present a related problem: when we have some prior information about any of the elements of the mixing matrix, and how traditional solutions can be modified incorporating this information to obtain new estimators.<sup>1</sup>

## 1. INTRODUCTION

Blind Source Separation (BSS) for instantaneous and noiseless mixture consists on recovering some statistically independent signals named *source signals* starting from linear mixtures of them. Recently a lot of papers in this area of signal processing have been published. This paper presents a modified problem related to BSS.

In BSS nothing is normally supposed about mixing matrix. If any prior knowledge is included in the statement of the problem, this is referred to the sources. For example, in [1] sources are temporally correlated, in [2], the probability density function (pdf) of the sources are known, and, in [3], sources are discrete.

With the aim of avoiding the inherent indetermination associated to the BSS problem (we do not know the order of the signals and their amplitudes) in some solutions some conditions are imposed to the mixing matrix or to the recovering matrix. In [4], the value of the elements of the diagonal are one, and we have to obtain only the off-diagonal elements. In [5], the modulus of the columns of the recovering matrix is one. Normally, in the solutions the indetermination is eliminated imposing restrictions in the statistical properties of the sources, as for example, assuming that the sources have unit variance.

In this paper the statement of the BSS problem is changed. We will introduce a new information in the problem: we will suppose that a prior information about some of the elements of the mixing matrix is given.

In Section 2 we will define how this information related to the elements of the mixing matrix can be mathematically formulated

from a statistical point of view and the consequences in traditional hypothesis of BSS.

In Section 3 we will obtain, starting from traditional BSS algorithms, a new modified class of them and show the explicit form of some of these modified algorithms. Finally some results of applying them are shown in Section 4.

For simplicity we will only suppose the real instantaneous case, although more examples are found in convolutive mixture (knowledge of some of the filters or, at least, kind of filters and some coefficients of them, that relate the sources with the observed signals).

## 2. PRIOR INFORMATION ABOUT MIXING MATRIX

There are two kinds of prior information about mixing matrix elements.

- **Deterministic case:** one or more elements of the mixing matrix are known. In this simple case, these elements are included directly in the solution (recovering matrix). We can find a simple example in convolutive sound mixtures; the microphones that record the sound are normally close to the sources, so the elements of the diagonal are usually considered to be one.
- **Not deterministic case:** we have a prior information about some elements but we do not know exactly their value. This degree of uncertainty can be statistically modeled by a (pdf). If this pdf of the mixing matrix elements is correct and meaningful, it is clear that it could be included in BSS formulation achieving better results. It must be clear that we are interested in maintaining the BSS statement of the problem, including a Bayesian perspective by considering that the pdf is a prior pdf of the mixing matrix elements, not like in traditional array signal processing where the matrix is more restricted.

We will call our problem AKICA (A priori Knowledge Independent Component Analysis) in order to clarify the notation and nomenclature, and to distinguish it from BSS or ICA problem.

<sup>1</sup> This paper is supported by Spanish Education Ministry under contract FEDER-CICYT 1230.



The AKICA, considered as an extension of BSS is mathematically formulated for the real, instantaneous, noiseless 2x2 case as:

$$\mathbf{y} = \mathbf{A}\mathbf{s} \quad (1)$$

where  $\mathbf{y}$  is the 2x1 observed signals vector,  $\mathbf{s}$  is the 2x1 source signals vector whose components are statistically independent and  $\mathbf{A}$  is the mixing matrix 2x2 with an associated matrix  $\mathbf{p}(\mathbf{A})$  whose element  $(i,j)$  is  $p(a_{ij})$ , the pdf of the element  $a_{ij}$  of the mixing matrix that represents the prior information.

This definition includes the deterministic case; if any of the elements of the mixing matrix is known, it can be considered as a random variable with a pdf expressed as a delta function allocated in the correct value. On the other side, if there is not any information about  $a_{ij}$ , a uniform pdf will be used.

Our assumptions about the sources will be the same as in BSS; statistical independence, no more than one Gaussian source, and, in order to eliminate the indetermination about the amplitude, unit variance.

### 3. MODIFIED ALGORITHMS INCLUDING PRIOR INFORMATION

Traditional solutions are based on the minimization-maximization of a function that measures the statistical independence of the recovered signals. As we do not know anything about sources (their pdf is unknown), a pure statistical analysis of the problem is difficult. In order to approximate the statistical independence of the sources, many approaches have been developed for BSS: [5] maximizes contrast functions derived from mutual information, [6] presents a algebraic solution based on joint diagonalisation of fourth order cumulant matrices, [7] is based on information theory...

We will include our prior information modifying the contrast functions that somehow measure the independence of the recovered signals. Most of BSS solutions employ a two-step method; first, the observed signals are decorrelated and normalized,

$$\mathbf{u} = \mathbf{L}^{-1}\mathbf{y}, \quad E\{\mathbf{u}\mathbf{u}^T\} = \mathbf{I} \quad (2)$$

and second the orthogonal matrix (a Givens rotation matrix function of the rotation angle) is calculated:

$$\mathbf{Q} = \frac{1}{\sqrt{1+\theta^2}} \begin{bmatrix} 1 & \theta \\ -\theta & 1 \end{bmatrix}, \quad \mathbf{s} = \mathbf{Q}^T \mathbf{u} \quad (3)$$

The first step is well studied in the bibliography (PCA analysis), so we will focus in the estimation of the Givens matrix angle. In [8] we present a more general and theoretical approach including the influence of the whitening step on the prior information matrix. In this paper we will focus on the results section.

We will show two examples based on two of the most important algorithms in BSS: EASI, an iterative solution [9], and Comon's solution as a batch-type algorithm [5].

Cardoso's algorithm EASI is based on the minimization of the contrast function

$$f(\theta) = \sum E |s_i|^4 \quad (4)$$

for normalized decorrelated negative kurtosis sources, where  $\theta$  is the rotation angle defining the Givens rotation matrix. The modified EASI minimizes (with the same hypothesis) the function

$$\phi(\theta) = (E |s_1|^4 + E |s_2|^4) \cdot (-p(\theta)) \quad (5)$$

Applying a gradient method we obtain the new adaptation rule. It consists on a term like the EASI solution weighted by the value of the pdf for the estimated angle and a new term that tries to minimize  $-p(\theta)$  (or maximize  $p(\theta)$ ).

In Comon's solution the simplified contrast function maximizes the sum of the squared marginal cumulants of the recovered signals

$$\Psi(\mathbf{Q}) = \sum_{i=1}^2 \Gamma_{ii\dots i}^2 \quad (6)$$

where  $\Gamma$  are the cumulants of  $\mathbf{s}$  (normally cumulants of order 4 are considered). Our new estimator will maximize

$$\Psi(\mathbf{Q}) = \sum_{i=1}^2 \Gamma_{ii\dots i}^2 \cdot p(\theta) \quad (7)$$

In all these new algorithms the most important role is played by the prior pdf. If it is not correct, probably our estimator will be biased. If the pdf is meaningful, the new estimator will have less variance.

However, there are BSS solutions that are not obtained starting from an objective (contrast) function. These algorithms are tests of statistical independence of the recovered signals in convergence. As an example, we consider Jutten-Herault learning rule [4]. In this case, our prior knowledge can be directly introduced adding a term that tries to adjust the solution in order to maximize the prior probability. The modified Jutten-Herault algorithm is:

$$c_{ij}[n+1] = c_{ij}[n] + \alpha \cdot f(s_i) \cdot g(s_j) + \beta \cdot \left. \frac{\partial p(c_{ij})}{\partial c_{ij}} \right|_{c_{ij}=c_{ij}[n]} \quad (8)$$

where the elements of the recovering matrix are  $c_{ij}$ ,  $\alpha, \beta$  are learning steps,  $f$  and  $g$  different odd functions. In convergence,

$$E\{s_i^{2j+1} \cdot s_k^{2m+1}\} + \frac{\partial p(c_{ij})}{\partial c_{ij}} = 0 \quad (9)$$

In (9) is clear that an accurate mathematical model of the prior knowledge is necessary if we want to obtain independent recovered signals.

## 4. RESULTS

### Example 1. EASI and modified EASI algorithms.

Source 1 is a sinusoidal function and source 2 a sawtooth function, both of them clearly subgaussian. They are mixed by a Givens matrix with  $\theta = 0.2$ .

In figure 1 we show the estimated angle for the EASI and modified EASI algorithms, both with the same initial condition and the same adaptation coefficient.

The prior information corresponds to a gaussian pdf, with mean 0.2 and unit variance. As we can see in this figure, both of them converge to the correct solution in mean, but the variance of the EASI solution is greater than in the new algorithm. If we want to reduce the variance we can reduce the adaptation step, but the convergence slows, so we need more samples to obtain the correct angle.

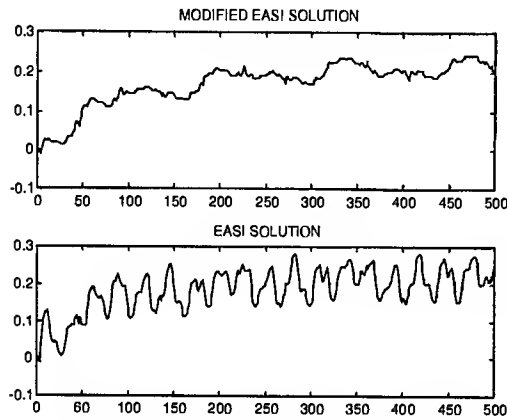


Figure 1. EASI and modified EASI solutions.

### Example 2. Comon and modified Comon algorithms.

It represents a mixture of two different sinusoidal signals. In this case  $\hat{\theta} = -0.5$ .

Figure 2 shows the estimated  $\theta$  in front of the number of observations; in dotted line the new estimator, and in solid line Comon's solution. The new algorithm converges faster than classical one. The prior pdf is a Gaussian with mean  $-0.5$  and unit variance.

Figure 3 compares the variance of both estimators. In this case, the modified estimator has less variance than the other one, and for 60 or more observations, both of them achieve the correct solution with a low variance. We can see that modified algorithm only need 40 samples to obtain an unbiased and low variance estimator for this example.

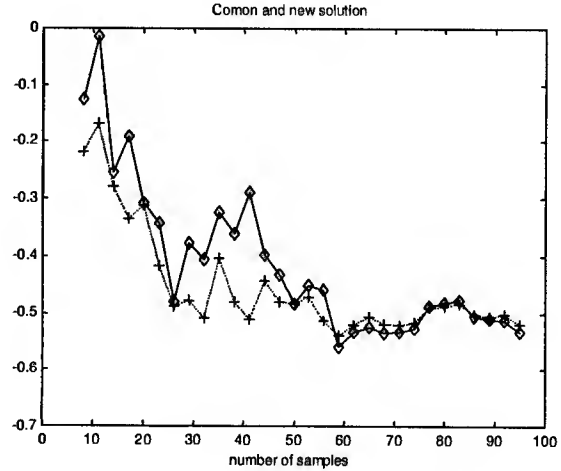


Figure 2. Estimated angle vs number of observations. In solid line Comon's solution and in dotted line the modified Comon's solution.

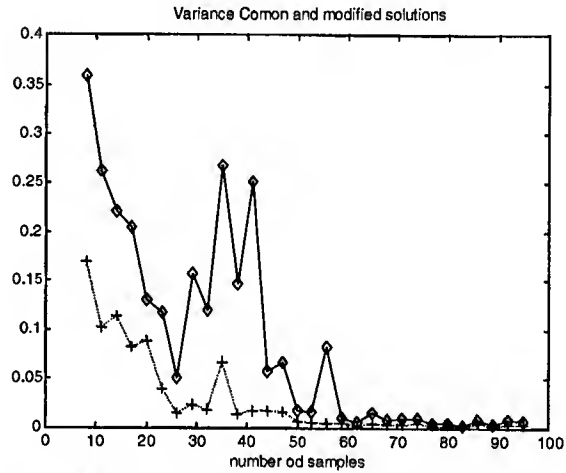
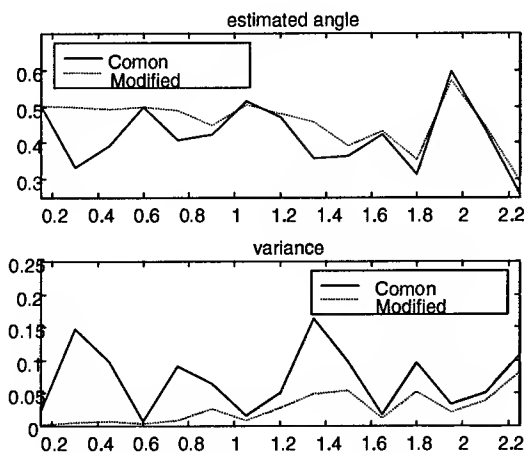


Figure 3. Variance of Comon (solid line) and modified Comon (dotted line) estimators vs number of observations.

### Example 3. Influence of prior knowledge (variance).

Correct angle is  $\hat{\theta} = 0.5$  and  $p(\theta)$  represents a Gaussian r.v. with mean 0.5 and variable variance. Only 30 samples are considered.

In figure 4 we observe that Comon's algorithm needs more observations to obtain the correct solution and to decrease the variance, while modified solution, for a prior knowledge with low variance, estimates the angle appropriately. However, when our prior knowledge is not significant (high variance), both of them are the same algorithms, and the solutions are similar.



**Figure 4.** Estimated angle (up) and variance of the estimator (down) of Comon (solid line) and modified Comon (dotted line) estimators vs variance of prior unbiased pdf.

#### Example 4. Influence of prior knowledge (mean).

The angle is 0.3 and prior pdf is modeled by a Gaussian r.v. with variable mean and variance 0.6. The number of observations is 50. In **figure 5** we show how if prior knowledge is biased and variance is low, the angle estimated by modified algorithm is biased, not Comon's solution. However, the variance of modified algorithm is lower than Comon's algorithm, so for incorrect prior information (mean far away from the correct mixing angle) with low variance, mean squared error is similar to traditional estimator.

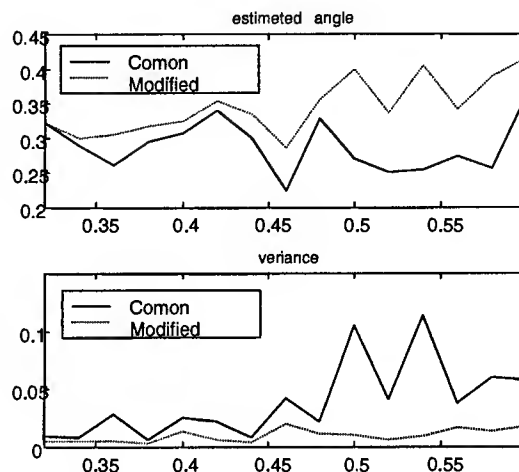
#### Example 5. Jutten-Herault and modified algorithms.

Source signals are the same that in example 1. The value of the coefficient  $c_{12}$  is 0.3. A Gaussian r.v. models our prior information (mean is 0.3 and variance is variable). In **figure 6** we show the recovered coefficient. In the unit-variance case only 50 samples are needed to estimate properly the mixing coefficient.

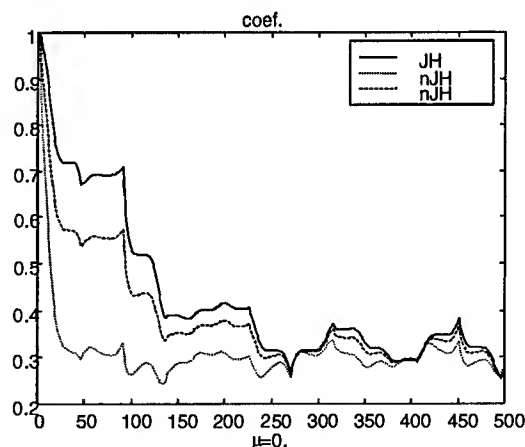
## 5. REFERENCES

- [1] A. Belouchrani, K. Abed-Meraim, J.F. Cardoso, E. Moulines, "Second order blind separation of correlated sources", *Proc. Int. Conf. Digital Signal Processing*, Chipre, 1993, pp 346-351.
- [2] D.T. Pham, P. Garat, C. Jutten, "Separation of a mixture of independent sources through a maximum likelihood approach", *Proc. EUSIPCO 92*, 1992, pp 771-774.
- [3] A. Belouchrani, J.F. Cardoso, "Maximum likelihood source separation by the expectation-maximization technique: deterministic and stochastic implementation", *Proc. NOLTA 95*, pp 49-53.
- [4] C. Jutten, J. Herault, "Blind separation of sources, Part I: an adaptive algorithm based on neuromimetic architecture", *Signal Processing*, Vol. 24, No. 1, July 1991, pp 1-10.

- [5] P. Comon, "Independent component analysis, a new concept?", *Signal Processing*, Vol. 36, No. 3, April 1994, pp 287-314.
- [6] J.F. Cardoso, A. Souloumiac, "Blind beamforming for non gaussian signals", *Proc. Inst. Elec. Eng., pt. F*, Vol. 140, pp. 362-370, Dec. 1993.
- [7] A.J. Bell, T. J. Sejnowsky, "An information maximization approach to blind separation and blind deconvolution", *Neural computation*, 7, 1129-1159.
- [8] J. Igual, L. Vergara, "Prior information about mixing matrix in BSS formulation", *Proc. ICA 2000*.
- [9] B. Laheld, J.F. Cardoso, "Adaptive source separation with uniform performance", *Proc. EUSIPCO'94*, pp. 183-186.



**Figure 5.** Estimated angle (up) and variance of the estimator (down) of Comon (solid line) and modified Comon (dotted line) estimators vs mean of prior 0.6 variance pdf, for 50 observations.



**Figure 6.** Recovered  $c_{12}$  coefficient vs number of observations for Jutten-Herault algorithm (solid line) and an unbiased prior information with variable variance (dotted line, variance is one; dash-dotted line, variance is two).

# APPROXIMATE MAXIMUM LIKELIHOOD BLIND SOURCE SEPARATION WITH ARBITRARY SOURCE PDFS

Mounir Ghogho<sup>1\*</sup>, Ananthram Swami<sup>2</sup> and Tariq Durrani<sup>1\*</sup>

mounir@spd.eee.strath.ac.uk

a.swami@ieee.org

durrani@strath.ac.uk

<sup>1</sup>Department of Electronic and Electrical Engineering, University of Strathclyde, Glasgow G1 1XW, UK.

<sup>2</sup>Army Research Lab, AMSRL-IS-TA, Adelphi, MD 20783-1197, USA

## ABSTRACT

We present a quasi-maximum likelihood approach to blind source separation (BSS) which is based on approximating the source distributions by their truncated Edgeworth expansions. The paper focuses on the  $2 \times 2$  case, for which the problem is known to reduce to the estimation of a single rotation angle. Unlike existing maximum likelihood BSS techniques, the proposed algorithm is consistent for any source distribution, provided that the usual identifiability condition (at most one Gaussian source) is satisfied. Closed-form expressions are derived for the true CRB, for the CRB corresponding to the Edgeworth approximation, and for the large-sample variance of the proposed estimator. The proposed algorithm is compared with existing approaches via extensive simulations.

## 1. INTRODUCTION

Blind signal estimation aims to estimate unknown source signals from distorted and noise-contaminated observations without explicit knowledge of the transmission channel. In the case of multiple sources transmitted simultaneously, blind signal estimation requires, in general, multiple sensors or antennas (or other forms of diversity, such as time, frequency, code, etc.). Consider the scenario where  $L$  source signals are impinging on a set of  $M$  sensors. The transfer function of the antenna array is assumed to be linear and memoryless. The mixing matrix may be unknown due to several reasons: sensor locations are unknown (or not exactly known), or the array is uncalibrated; angular spread (due to local scattering, source signals impinge on the array at different angles but with no appreciable delay) also causes the array matrix to lose its usual Vandermonde structure. Thus, the outputs of the antenna array are modelled as

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{v}(t) \quad (1)$$

where  $\mathbf{A} = [a_{ij}]$  is the  $(M \times L)$  mixing matrix characterizing the antenna array,  $\mathbf{s}(t)$  is the  $(L \times 1)$  source signal vector,  $\mathbf{v}(t)$  is the  $(M \times 1)$  additive noise vector, and  $t$  is the time index. The objective is to blindly recover or estimate the source vector,  $\mathbf{s}(t)$ . A less demanding objective is to recover  $\mathbf{s}(t)$  up to scale factors and permutation ambiguities. BSS consists of finding a separating matrix  $\mathbf{B}$  such that  $\mathbf{B}\mathbf{A}$  is a non-mixing matrix (see [4]), i.e.,  $\mathbf{B}\mathbf{A} = \mathbf{P}$  where  $\mathbf{P}$  is a generalized permutation matrix (generalized in the sense that the non-zero entries are not constrained to be unity). To ensure that the spatial signatures of the signals incident on the array are distinct, the mixing matrix must satisfy the following assumption.

(AS1)  $\mathbf{A}$  has full column rank.

In general, BSS requires the following extra assumptions:

(AS2)  $s_i(t)$ ,  $i = 1, \dots, L$ , are mutually independent.

(AS3)  $s_i(t)$  are zero-mean stationary and mixing.

(AS4)  $\mathbf{v}(t)$  is an  $M$ -variate zero-mean stationary mixing process and is independent of the source signals.

(AS5) At most one of the sources is Gaussian. More precisely, we assume that  $|\kappa_{4,1}| + |\kappa_{4,2}| \neq 0$ ; see eq. (4).

BSS techniques can be categorized into two classes: *i*) statistical distribution-based techniques [2] [4][7] which exploit *a priori* information about the statistical or deter-

ministic distribution of the source vector, such as, finite alphabet, non-Gaussianity, constant modulus, known skewness, kurtosis or other statistics; *ii*) temporal-correlation-based techniques [10] [1] under the assumption that the source signals have distinct (not necessarily orthogonal) time-frequency signatures. The BSS problem has attracted increasing attention because of its wide range of applications, such as biomedical engineering, radar processing, acoustic tracking and mobile communications.

A standard approach to BSS is to first spatially pre-whiten the data using second-order statistics; this serves to normalize the signal powers as well, and reduces the mixing matrix to a unitary matrix. In the second step, this unitary matrix is estimated. Here, we focus on the noiseless (high SNR) case and assume that the pre-whitening step has been carried out perfectly. The discrete-time data vector is now given by

$$\mathbf{z}(t) = \mathbf{U}\mathbf{s}(t), \quad t = 1, 2, \dots \quad (2)$$

where  $\mathbf{U}$  is the unknown unitary or rotation matrix. The covariance matrix of  $\mathbf{z}(t)$  is the identity matrix. We also focus on the basic scenario of two sources-two sensors. As explained in [4], this scenario can be used to solve the general problem by operating pairwise over several sweeps until convergence. The unknown  $(2 \times 2)$  unitary transformation matrix is reduced to a Givens rotation matrix, which can be expressed in the case of real-valued signals as

$$\mathbf{U}(\theta_o) = \begin{pmatrix} \cos(\theta_o) & -\sin(\theta_o) \\ \sin(\theta_o) & \cos(\theta_o) \end{pmatrix} \quad (3)$$

where  $\theta_o$  is the rotation angle between the two source signals. The source vector  $\mathbf{s}(t)$  can be unambiguously estimated if  $\theta_o$  can be unambiguously identified. However, in order for BSS to be achieved, only  $[\theta_o]_{\pi/4}$  needs to be identified, where  $[\theta_o]_{\alpha}$  denotes the contracted value of  $\theta$  in the interval  $[-\alpha, \alpha)$ . Indeed,  $\mathbf{U}([\theta_o]_{\pi/4})\mathbf{U}(\theta_o)$  is a non-mixing matrix. With  $[\theta_o]_{\pi/4}$ , the signal vector  $\mathbf{s}(t)$  can be identified only up to sign and permutation ambiguities. The permutation ambiguity can be relieved if  $[\theta_o]_{\pi/2}$  can be identified, which requires the source signals to have different distributions. The sign ambiguity can be relieved if the source distributions are non-symmetric.

An approximate maximum likelihood BSS was proposed in [6] where the Gram-Charlier expansion was used to approximate the source distributions. The algorithm proposed in [6] is limited to the case where the sum of the source kurtoses is positive, i.e.,  $\kappa_{4,1} + \kappa_{4,2} > 0$ , where

$$\kappa_{4,i} = E(s_i^4(t))/[E(s_i^2(t))]^2 - 3, \quad (4)$$

is the kurtosis of the  $i$ th source. An extension of this algorithm to include the  $\kappa_{4,1} + \kappa_{4,2} < 0$  case was suggested in [11] using a geometrical approach. However, both of these methods fail when  $\kappa_{4,1} + \kappa_{4,2}$  is close to zero. The closed-form BSS solution proposed by Comon [3, 4], which is based on the concept of independent component analysis (ICA) also fails in this scenario. In this paper, we generalize these methods in order to overcome the above problem. The proposed algorithms are consistent for any source distributions provided that the identifiability condition (which

\*Partly supported by the Engineering and Physical Sciences Research Council of UK under research grant ref. GR/L07475.

is: at most one of the sources is Gaussian [10][4]; see (AS5)) is satisfied.

The scenario where  $\kappa_{4,1} + \kappa_{4,2}$  is close to zero can be encountered when an impulsive signal interferes with a communication signal. Indeed, the kurtosis of the latter is negative-valued whereas that of an impulsive interference is positive-valued. The proposed algorithm will be shown to be *robust* in this scenario, and yet its performance in the absence of impulsive interference is comparable with that of existing BSS techniques.

We derive a closed-form expressions for: (1) the exact CRB; (2) the CRB corresponding to the truncated Gram-Charlier series; and (3) the large sample variance of our estimator. These expressions are also applicable to the estimates of the rotation angle of a complex symbol set with independent real and imaginary parts.

## 2. AN APPROXIMATE MAXIMUM LIKELIHOOD (AML) APPROACH

Let the normalized (wrt the unknown scale) pdf of the  $i$ th source be denoted by  $p_{s_i}(\cdot)$ . Since the sources are independent, and ignoring the whitening imperfections, the ML estimate of the rotation angle  $\theta_o$  is obtained by maximizing

$$l(Z_T; \theta) = \hat{E}_T \left\{ \sum_{i=1,2} \ln p_{s_i}([U(\theta)^T z]_i) \right\}$$

where  $Z_T = [z(1), \dots, z(T)]$ ,  $z(t)$  is given by (2) and (3),  $\hat{E}_T\{\mathbf{y}\} = 1/T \sum_{t=1}^T \mathbf{y}(t)$  and  $[y]_i$  denotes the  $i$ th entry of vector  $\mathbf{y}$ . The ML scheme may be simplified by approximating the pdf of the sources by their truncated Edgeworth or Gram-Charlier expansions [8], which involve cumulants. Both of these expansions are ways to express a pdf by an expansion around the Gaussian kernel.

### 2.1. Approximate Likelihood Function

The Edgeworth expansion of the pdf of  $s_i$  about its best Gaussian approximate is [8]

$$p_{s_i} = g(s_i)[1 + \delta\varphi_i(s_i)]$$

where  $g(\cdot)$  is the normalized Gaussian distribution, and

$$\delta\varphi_i(s) = \frac{1}{3!}\kappa_{3,i}h_3(s) + \frac{1}{4!}\kappa_{4,i}h_4(s) + \dots$$

In the above expression,  $\kappa_{n,i}$  is the  $n$ th-order cumulant of  $s_i$  and  $h_i(\cdot)$  is the Hermite polynomial of degree  $i$ . An approximation of the source pdf is obtained by truncating the Edgeworth expansion as follows

$$p_{s_i}(s_i) \approx g(s_i) \left[ 1 + \frac{\kappa_{3,i}}{3!}(s_i^3 - 3s_i) + \frac{\kappa_{4,i}}{4!}(s_i^4 - 6s_i^2 + 3) \right] \quad (5)$$

The same approximation is obtained via the Gram-Charlier expansion.

Using the polar coordinates of  $\mathbf{z}$ , i.e.,

$$\rho = |z_1^2 + z_2^2|^{1/2}, \quad \phi = \angle\{z_1 + jz_2\},$$

$$U(\theta)^{-1} \mathbf{z} = \begin{pmatrix} \rho \cos(\phi - \theta) \\ \rho \sin(\phi - \theta) \end{pmatrix}. \quad (6)$$

Using (5) and (6), the likelihood function can be approximated as (using  $\ln(1+c) \approx c$ , for  $|c| \ll 1$ , and dropping non-relevant terms)

$$l(Z_T; \theta) = \hat{E}_T\{l_3(\mathbf{z}; \theta)\} + \hat{E}_T\{l_4(\mathbf{z}; \theta)\} \quad (7)$$

where

$$\begin{aligned} l_3(\mathbf{z}; \theta) &= \frac{\rho^3}{8} [\kappa_{3,1} \cos(\phi - \theta) + \kappa_{3,2} \sin(\phi - \theta)] \\ &\quad + \frac{\rho^3}{24} [\kappa_{3,1} \cos 3(\phi - \theta) - \kappa_{3,2} \sin 3(\phi - \theta)], \\ l_4(\mathbf{z}; \theta) &= \frac{\kappa_{4,1} - \kappa_{4,2}}{48} \rho^4 \cos[2(\phi - \theta)] \\ &\quad + \frac{\kappa_{4,1} + \kappa_{4,2}}{192} \rho^4 \cos[4(\phi - \theta)]. \end{aligned}$$

### 2.2. Estimation

The AML estimator of  $\theta_o$  is obtained by maximizing  $\hat{E}_T\{l(Z_T; \theta)\}$ . We have that

$$\theta_o = \arg \max_{\theta} E\{l_3(\mathbf{z}; \theta)\},$$

$$\theta_o = \arg \max_{\theta} E\{l_4(\mathbf{z}; \theta)\}.$$

The estimation of  $\theta_o$  using  $l_3(\mathbf{z}; \theta)$  is feasible only if at least one of the two sources has non-zero skewness, which may be restrictive in practice. But  $\theta_o$  can be consistently estimated by maximizing  $\hat{E}_T\{l_4(\mathbf{z}; \theta)\}$  regardless of the pdf of the source signals; we stress that this holds true without invoking the assumption of symmetric pdfs, as in [6].

Notice that  $l_4(\mathbf{z}; \theta)$  consists of two terms, one is a function of the difference of the source kurtoses and the other is a function of the sum of these kurtoses. Let

$$l_4^-(\mathbf{z}; \theta) = \frac{\kappa_{4,1} - \kappa_{4,2}}{48} \rho^4 \cos[2(\phi - \theta)],$$

$$l_4^+(\mathbf{z}; \theta) = \frac{\kappa_{4,1} + \kappa_{4,2}}{192} \rho^4 \cos[4(\phi - \theta)].$$

Interestingly, these functions are also 'likelihoods' (or contrasts [4]) for the estimation of  $\theta_o$ , i.e.,

$$\theta_o = \arg \max_{\theta} E\{l_4^-(\mathbf{z}; \theta)\}, \quad (8)$$

$$\theta_o = \arg \max_{\theta} E\{l_4^+(\mathbf{z}; \theta)\}. \quad (9)$$

Let  $\hat{\theta}_o^-$  and  $\hat{\theta}_o^+$  denote the corresponding estimators. First notice that  $\hat{\theta}_o^-$  is consistent if  $\kappa_{4,1} - \kappa_{4,2} \neq 0$ , and  $\hat{\theta}_o^+$  is consistent if  $\kappa_{4,1} + \kappa_{4,2} \neq 0$ . Since these conditions cannot be violated simultaneously under assumption (AS5), at least one of these estimators is consistent. To derive these estimators, we first notice that

$$l_4^-(\mathbf{z}; \theta + \pi) = l_4^-(\mathbf{z}; \theta),$$

$$l_4^+(\mathbf{z}; \theta + \pi/2) = l_4^+(\mathbf{z}; \theta).$$

This implies that the identifiability range is  $[-\pi/2, \pi/2]$  for  $\hat{\theta}_o^-$  and  $[-\pi/4, \pi/4]$  for  $\hat{\theta}_o^+$ . Therefore,  $\hat{\theta}_o^-$  is actually an estimate of  $[\theta_o]_{\pi/2}$ , and  $\hat{\theta}_o^+$  is an estimate of  $[\theta_o]_{\pi/4}$ . The identifiability range can be further extended by using  $l_3(\mathbf{z}; \theta)$ , if the sources are not both symmetrically distributed.

After some algebra, we obtain the following closed-form estimators

$$\hat{\theta}_o^- = \frac{1}{2} \arg \{ \text{sign}(\kappa_{4,1} - \kappa_{4,2})(\eta_1 + j\eta_2) \}, \quad (10)$$

$$\hat{\theta}_o^+ = \frac{1}{4} \arg \{ \text{sign}(\kappa_{4,1} + \kappa_{4,2})(\eta_3 + j\eta_4) \}, \quad (11)$$

where

$$\eta_1 = \hat{E}_T \{ \rho^4 \cos 2\phi \}; \quad \eta_2 = \hat{E}_T \{ \rho^4 \sin 2\phi \}$$

$$\eta_3 = \hat{E}_T \{ \rho^4 \cos 4\phi \}; \quad \eta_4 = \hat{E}_T \{ \rho^4 \sin 4\phi \}.$$

Note that the ML estimator proposed in [6], i.e.,  $\arg\{(\eta_3 + j\eta_4)/4\}$ , coincides with  $\hat{\theta}_o^+$  when  $\kappa_{4,1} + \kappa_{4,2} > 0$  (in [6], the kurtoses of the sources were assumed positive and equal; we note that this assumption is not valid in the communications context).

For the above estimators to be practical, the signs of  $(\kappa_{4,1} - \kappa_{4,2})$  and  $(\kappa_{4,1} + \kappa_{4,2})$  must be known or estimated from the data. After some calculations, we obtain

$$\begin{aligned}\kappa_{4,1} - \kappa_{4,2} &= E\{\rho^4 \cos 2\phi\} \cos 2\theta_o + E\{\rho^4 \sin 2\phi\} \sin 2\theta_o, \\ \kappa_{4,1} + \kappa_{4,2} &= E\{\rho^4\} - 8.\end{aligned}$$

Hence, the latter can be estimated directly from the data. The former however depends on the unknown rotation angle  $\theta_o$ . A *completely* blind version of the estimator  $\hat{\theta}_1^+$  is then obtained as

$$\hat{\theta}_1^+ = \frac{1}{4} \arg\left\{\text{sign}\left(\hat{E}_T\{\rho^4\} - 8\right)(\eta_3 + j\eta_4)\right\}. \quad (12)$$

This estimator was also proposed in [11] using a geometrical approach. Since  $\hat{\theta}_1^+$  fails when  $\kappa_{4,1} + \kappa_{4,2}$  is close to 0, we need to develop a general estimator that would be consistent regardless of the source distributions. Towards this objective, we go back to the likelihood  $\hat{E}_T\{l_4(z; \theta)\}$  in eq. (7) and replace  $\kappa_{4,1} - \kappa_{4,2}$  and  $\kappa_{4,1} + \kappa_{4,2}$  by their estimates

$$\begin{aligned}\widehat{\kappa_{4,1} - \kappa_{4,2}} &= \eta_1 \cos 2\theta + \eta_2 \sin 2\theta, \\ \widehat{\kappa_{4,1} + \kappa_{4,2}} &= \zeta := \hat{E}_T\{\rho^4\} - 8,\end{aligned}$$

After some tedious calculations, the LLF is found to be

$$\begin{aligned}\hat{E}_T\{l_4(Z_T; \theta)\} &= \frac{1}{96}(\eta_1^2 + \eta_2^2) \\ &\quad - \frac{1}{192}[2\eta_1^2 - 2\eta_2^2 + \zeta\eta_3] \cos 4\theta \\ &\quad + \frac{1}{192}[4\eta_1\eta_2 + \zeta\eta_4] \sin 4\theta. \quad (13)\end{aligned}$$

Let  $\hat{\theta}_o$  denote the maximizer of  $\hat{E}_T\{l_4(Z_T; \theta)\}$ . Setting the first derivative of the LLF in eq. (13) to 0, the AML estimate of  $\theta_o$  is then obtained as

$$\hat{\theta}_o = \frac{1}{4} \arg\{e_1 + je_2\} \quad (14)$$

where

$$e_1 = 2\eta_1^2 - 2\eta_2^2 + \zeta\eta_3; \quad e_2 = 4\eta_1\eta_2 + \zeta\eta_4.$$

Notice that after replacing  $\kappa_{4,1} - \kappa_{4,2}$  by its estimate, the  $2\theta$  terms in the LLF have disappeared, and the LLF becomes a function of  $4\theta$  only. This implies that the above AML estimate  $\hat{\theta}_o$  is an estimate of  $[\theta_o]_{\pi/4}$ . However,  $[\theta_o]_{\pi/2}$  can be estimated if  $\kappa_{4,1} - \kappa_{4,2} \neq 0$ . Indeed, if the source signals are ordered such that  $\kappa_{4,1} > \kappa_{4,2}$ , the AMLE of  $[\theta_o]_{\pi/2}$  is  $\hat{\theta}_o$ , if  $\eta_1 \cos 2\hat{\theta}_o > -\eta_2 \sin 2\hat{\theta}_o$ , and is  $[\hat{\theta}_o + \pi]_{\pi/2}$  otherwise.

### 2.3. Summary of the AML Algorithm

(1) Compute the whitening matrix,  $\mathcal{W}$ ; whiten the array outputs and normalize their powers:

$$\mathbf{z} = \mathcal{W}^{-1}\mathbf{x} \quad (=U\mathbf{s})$$

(2) Evaluate  $\eta_i$ ,  $i = 1, \dots, 4$ , via

$$\begin{aligned}\eta_1 &= \hat{E}_T\{z_1^4 - z_2^4\}; & \eta_2 &= 2\hat{E}_T\{z_1^3 z_2 + z_1 z_2^3\} \\ \eta_3 &= \hat{E}_T\{z_1^4 - 6z_1^2 z_2^2 + z_2^4\}; & \eta_4 &= 4\hat{E}_T\{z_1^3 z_2 - z_1 z_2^3\}\end{aligned}$$

(3) Estimate the sum of the source kurtoses via

$$\zeta = \hat{E}_T\{(z_1^2 + z_2^2)^2\} - 8$$

(4) Estimate  $\theta$  via eqn. (14),

$$\hat{\theta}_o = \frac{1}{4} \text{atan2}(4\eta_1\eta_2 + \zeta\eta_4, 2\eta_1^2 - 2\eta_2^2 + \zeta\eta_3).$$

### 3. PERFORMANCE ANALYSIS

For simplicity, we assume in the rest of the paper that the source signals are temporally independent; such an assumption is not required to ensure the consistency of the algorithms developed in preceding sections (it suffices to assume that the source signals are weakly mixing in the sense that they have summable cumulants).

We have derived a closed-form expression for the large sample variance of the completely blind estimator given by (12), which assumes  $\kappa_{4,1} + \kappa_{4,2} \neq 0$ . The expression for the case where  $\kappa_{4,1} - \kappa_{4,2} \neq 0$  is considerably more complicated, and will be presented elsewhere.

The large-sample variance expression is given by

$$\text{var}(\hat{\theta}) = \frac{1}{T} \frac{Es_1^6 - 2Es_1^4 Es_2^4 + Es_2^6}{(\kappa_{4,1} + \kappa_{4,2})^2} \quad (15)$$

Details of the derivation are omitted due to lack of space and may be found in [5].

**Remarks:**

1. Derivation assumes that  $s_i(t)$ 's are zero-mean, unit variance iid sequences (valid for BSS), with  $\kappa_{4,1} + \kappa_{4,2} \neq 0$ .
2. The estimator involves the fourth moments of  $z(\cdot)$ , but the variance expression does not depend upon the eighth order moments of the source signals.
3. The expression holds true even if neither source pdf is symmetric.

### 4. CRAMÉR-RAO BOUNDS

We derive the CRB for  $\theta_o$  when the estimator is based on the LLF in eq. (7) corresponding to the Gram-Charlier expansion. We also derive the true CRB (i.e., using the exact pdf). Detailed derivations may be found in [5].

#### 4.1. True or Exact CRBs

We show that :

- (i) The CRB for  $\theta_o$  and the CRB for the source pdf parameters are decoupled;
- (ii) Under mild assumptions on the pdfs, the true or exact CRB is given by,

$$\text{CRB}(\theta_o) = \frac{1}{N} [I_{p_1} + I_{p_2} - 2]^{-1}$$

where  $I_p$  is the Fisher information for location (FIL) of the (standardized) pdf  $p(\cdot)$ , and is defined as

$$I_p = \int_{-\infty}^{\infty} [p'(s)]^2 / p(s) ds.$$

We next consider a few special cases:

- (a) For the Gaussian pdf,  $I_p = 1$ ; hence, if both sources are Gaussian,  $\theta$  cannot be estimated, as is well known.
- (b) Generalized Gaussian pdf with shape parameter  $\alpha$ :

$$I_p = \alpha^2, (2 - 1/\alpha), (3/\alpha) / \alpha^2, (1/\alpha).$$

Note that  $I_p(\alpha_1) + I_p(\alpha_2) = 2$  iff  $\alpha_1 = \alpha_2 = 2$ , i.e., the Gaussian case. If  $\alpha = 1$  (Laplace), we have  $I_p = 2$ .

#### 4.2. CRB for the Gram-Charlier Approximation

We show that

(i) Assuming that the skewness and kurtoses are known,

$$CRB_{\theta}(\theta_o) = \frac{1}{N} \left[ \frac{1}{2}(\kappa_{3,1}^2 + \kappa_{3,2}^2) + \frac{1}{6}(\kappa_{4,1}^2 + \kappa_{4,2}^2) \right]^{-1} \quad (16)$$

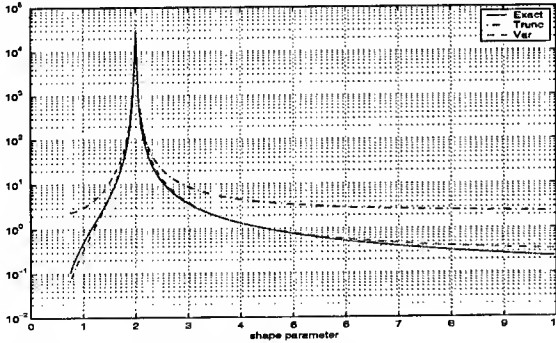
(ii) If  $\kappa_{3,i}$  or  $\kappa_{4,i}$  are not known, the FIM for these parameters is identically zero, and the CRB for  $\theta_o$  (computed via the pseudo-inverse) remains the same as in (i).

**Remarks:**

(a) It is surprising that the FIM for the kurtosis is zero; but, once we find  $\theta$ , we can derotate and estimate the two source signals, from which the skew/kurtosis can be estimated; however, since the two sources may be swapped, the estimated kurtoses also may be swapped. Thus, FIM = 0 is due entirely to the permutation ambiguity. In the scalar case, given  $x(t) = as(t)$ , where  $s(t)$  is unit variance, iid, one can estimate both  $a$ , and the kurtosis of  $s$  consistently.

(b) When the sources are symmetrically distributed, the third-order cumulants vanish and the CRB reduces to that given in [6]. The CRB is symmetric in the skewnesses and kurtoses of the two sources. If both sources have zero skewness and zero kurtoses, then the CRB is infinity, and  $\theta$  cannot be estimated, since the truncated Gram-Charlier expansion reduces to the Gaussian pdf. It is also interesting that the CRB expression does not involve the sign of the skewness or the kurtosis.

(c) Both CR bounds are independent of the true value of  $\theta$ . In order to compare them, we consider the case where the source signals are generalized-Gaussian,  $p(s) \propto \exp(-|s|^a)$ . Figure (1) shows the 2 CRBs for the case where both source signals have the same shape parameter. Notice that in the heavy-tailed case ( $a < 2$ ), the exact and truncated bounds are close. In the lighter-tailed case ( $a > 2$ ), the CRB corresponding to the truncated estimator is quite pessimistic.

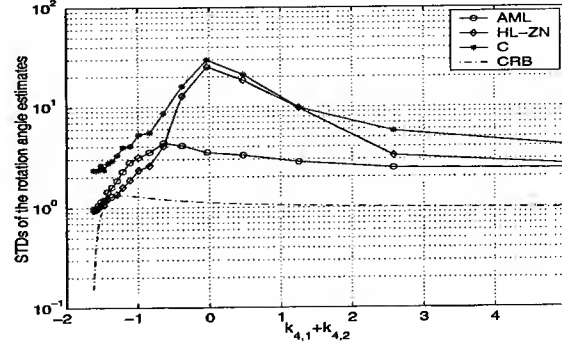


**Figure 1.** Exact and approximate CRBs and large sample variance of the estimator in (12) vs. shape parameter  $a$  of the generalized Gaussian pdf; both sources had the same pdf.  $T = 1$ .

#### 5. SIMULATION RESULTS

An extensive set of simulations were carried out to compare the five methods: (1) AML - method proposed in Section II of this paper; (2) C - method in [4]; (3) HL - method in [6]; (4) SGS - based on a result by Swami *et al* in [9] and, (5) ZN - method in [11]

A fixed rotation matrix corresponding to  $\theta_o = 15^\circ$  was chosen, and the number of samples was  $N = 5000$ . The input source signals were iid sequences drawn from different pdfs. Simulation results are summarized from  $K = 1000$  runs. Table 1 shows the results of the simulation study for 3 different examples (corresponding to different source pdfs). For each example, the four rows show the mean,



**Figure 2.** Standard deviation of the rotation angle estimates vs.  $\kappa_{4,1} + \kappa_{4,2}$  with  $\kappa_{4,1} = -0.8$ ;  $N=5000$ ;  $SNR = \infty$ . Both sources had the generalized Gaussian pdf.

AML	C	HL	SGS	ZN
Uniform and Laplace pdfs				
0.0480	-1.0267	0.0549	0.8902	0.1499
2.1566	5.6890	3.7684	4.7752	4.0493
-6.0813	-25.8391	-12.0107	-12.6000	-46.8333
6.2153	13.3698	11.0330	25.0000	11.0330
Uniform and Exponential pdfs				
-0.1126	-0.5541	-0.1105	0.1719	-0.1105
2.0792	3.2111	2.6050	2.9175	2.6050
-5.7475	-13.0922	-7.8230	-7.7500	-7.8230
5.8139	7.1190	7.5241	12.0000	7.5241

**Table 1.** Comparison of different estimators

standard deviation, minimum and maximum of the bias of the estimate. The source signals were changed randomly from realization to realization.

Another simulation example is reported in Figure 2.

#### 6. EXTENSIONS

In this section, we generalize our AML estimator so that the likelihoods  $l_4^-(z; \theta)$  and  $l_4^+(z; \theta)$  can be combined arbitrarily. Furthermore, we extend Comon's estimator [4], which is based on the concept of ICA, to the  $\kappa_{4,1} + \kappa_{4,2} = 0$  case.

##### 6.1. Weighted AML (WAML) estimator

Consider the more general estimator

$$\hat{\theta}_o(w) = \frac{1}{4} \text{atan2}(w(4\eta_1\eta_2) + (1-w)\zeta\eta_4, w(2\eta_1^2 - 2\eta_2^2) + (1-w)\zeta\eta_3) \quad (17)$$

where  $w$  ( $0 \leq w \leq 1$ ) is a weight parameter. This estimator reduces to that in eq. (12) when  $w = 0$ , and to that in eq. (14) when  $w = 0.5$ . This latter case is obtained by equally combining  $l_4^-(z; \theta)$  and  $l_4^+(z; \theta)$ . The former case is obtained by ignoring  $l_4^-(z; \theta)$ . The weight parameter,  $w$ , could be adjusted using *a priori* information about the source pdfs.

##### 6.2. Generalized ICA-based BSS

With  $z(t)$  and  $U$  as defined in (2) and (3), we can write the cross-cumulants of the output in terms of input cumulants defined in (4) and the unknown angle  $\theta$ . There are five distinct fourth-order cumulants involving two sensors

$$\begin{aligned} c_{1111} &= \kappa_{4,1} \cos^4(\theta_o) + \kappa_{4,2} \sin^4(\theta_o) \\ c_{2222} &= \kappa_{4,1} \sin^4(\theta_o) + \kappa_{4,2} \cos^4(\theta_o) \\ c_{1112} &= \kappa_{4,1} \cos^3(\theta_o) \sin(\theta_o) - \kappa_{4,2} \cos(\theta_o) \sin^3(\theta_o) \end{aligned}$$

$$c_{2221} = \kappa_{4,1} \cos(\theta_o) \sin^3(\theta_o) - \kappa_{4,2} \cos^3(\theta_o) \sin(\theta_o)$$

$$c_{1122} = (\kappa_{4,1} + \kappa_{4,2}) \cos^2(\theta_o) \sin^2(\theta_o)$$

where  $c_{ijkl} = \text{cum}(z_i, z_j, z_k, z_l)$ . Using some of these equations, we suggest the following estimator

$$\tan \tilde{\theta}_o = -\frac{\beta}{2} + \text{sign}(\beta) \sqrt{\frac{\beta^2}{4} + 1}. \quad (18)$$

where

$$\beta = \frac{w(c_{1111} - c_{2222}) + (1-w)(c_{1112} - c_{1222})}{w(c_{1112} + c_{1222}) + (1-w)c_{1122}}. \quad (19)$$

Note that  $c_{1111} - c_{2222}$  and  $c_{1112} + c_{1222}$  are proportional to  $\kappa_{4,1} - \kappa_{4,2}$ , and  $c_{1112} - c_{1222}$  and  $c_{1122}$  are proportional to  $\kappa_{4,1} + \kappa_{4,2}$ . When  $w = 0$ , the estimator reduces to Comon's estimator [3], which is consistent only if  $\kappa_{4,1} + \kappa_{4,2} \neq 0$ .

### 6.3. Simulation results

The source signals were generated using the generalized Gaussian pdf, and  $N = 5000$ . Figure 3 displays the standard deviations (STDs) of the estimates in eqs. (17) and (18) vs.  $w$  when the source pdfs are identical (uniform pdfs). It is seen that the proposed WAML estimator is less sensitive to  $w$  than the proposed ICA-based estimator. The WAML estimator with  $w = 0.5$  seems to perform almost as good as the optimal WAML estimator (which is obtained for  $w = 0$  in this case), especially when the source kurtoses are negative valued.

Figure 4 displays the STDs of the angle estimates when  $\kappa_{4,1} + \kappa_{4,2} = -0.06$  (with  $\kappa_{4,1} = 0.92$ ). As expected, both estimates fail when  $w = 0$ , and the optimal estimates are obtained when  $w = 1$ . It is also seen that the ICA-based estimator with a good choice of  $w$  can outperform the WAML estimator in some scenarios.

Figure 5 displays the STDs of the angle estimates when  $\kappa_{4,1} = -1$ ,  $\kappa_{4,2} = 0$ . In this case, the optimal value of  $w$  lies between 0 and 1.

Note however that by choosing  $0 < w < 1$ , only the accuracy of the estimates is affected, in contrast with the extreme cases,  $w = 0$  and  $w = 1$ , where the estimates may completely fail. This suggests that even if the source signals are known to have the same distribution (e.g., communication signals), it might be beneficial to choose  $w > 0$  (e.g.,  $w = 0.2$ ) in order to make the BSS robust to impulsive interference.

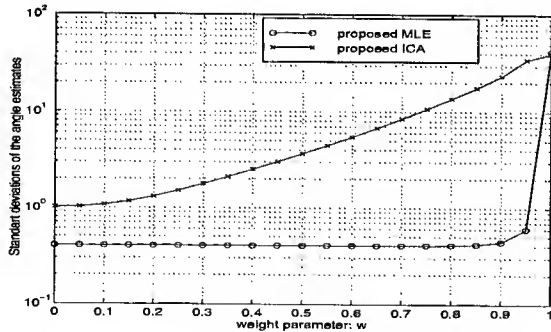


Figure 3. STD of the rotation angle estimates when  $\kappa_{4,1} = \kappa_{4,2} = -1$ .

### 6.4. Extensions

Future work includes the extension of the proposed BSS techniques to complex valued signals, and performance analysis. The latter will be useful to estimate the optimal value of the weight parameter  $w$  in (17) and (18), from the array outputs.

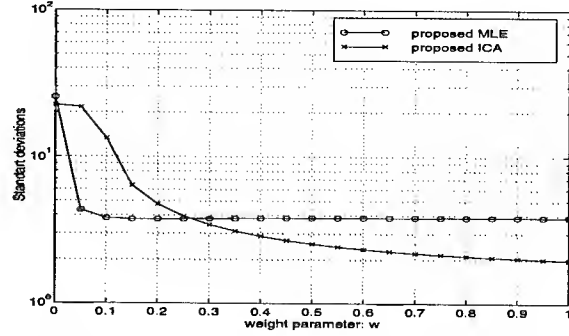


Figure 4. STD of the rotation angle estimates when  $\kappa_{4,1} + \kappa_{4,2} = -0.06$ ,  $\kappa_{4,1} = 0.92$

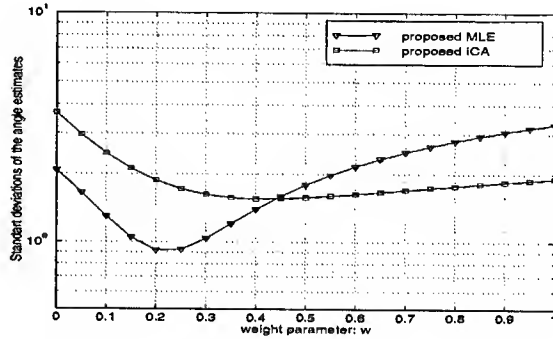


Figure 5. STD of the rotation angle estimates when  $\kappa_{4,1} = -1$ ,  $\kappa_{4,2} = 0$ .

## REFERENCES

- [1] A. Belouchrani and M.G. Amin, "Blind source separation based upon time-frequency signal representations", *IEEE Trans. Sig. Proc.*, 46(11), 2888-97, 1998.
- [2] J.-F. Cardoso, "Blind Signal Separation: Statistical Principles", *Proc. IEEE*, 86(10), 2009-25, Oct 1998.
- [3] P. Comon, "Separation of sources using higher-order cumulants", *Proc. SPIE*, vol 1152, 170-181, San Diego, 1989.
- [4] P. Comon, "Independent component analysis, a new concept?", *Signal Processing*, 36(3), 287-314, 1996.
- [5] M. Ghogho, A. Swami and T. Durrani, "Approximate Maximum Likelihood Blind Source Separation with Arbitrary Source Pdfs: Cramer-Rao Bounds and Performance," submitted to *IEEE Trans. Sig. Proc.*
- [6] F. Harroy and J.L. Lacoume, "Maximum likelihood estimators and Cramer-Rao bounds in source separation," *Signal Processing*, 55(3), 167-177, Dec. 1996.
- [7] D.B. Hillis, B.M. Sadler, A. Swami, "Independent Component Analysis Using a Genetic Algorithm", to appear in *Proc. SPIE'00*, Orlando, FL, Apr 2000.
- [8] A. Stuart and J.K. ORD, *Kendall's advanced theory of statistics, vol. 1: Distribution theory*, Wiley, 1987.
- [9] A. Swami, G.B. Giannakis and S. Shamsunder, "Multichannel ARMA processes", *IEEE Trans. Sig. Proc.*, 42(4), 898-913, April 1994.
- [10] L. Tong, Y. Inouye, and R. Lui, "Waveform-preserving blind estimation of multiple independent sources," *IEEE Trans. Sig. Proc.*, 41(7), 2461-2470, July 1993.
- [11] V. Zarzoso and A. K. Nandi, "Blind separation of independent sources for virtually any source probability density function," *IEEE Trans. Sig. Proc.*, 47(9), 2419-2432, 1999.



# Power Spectral Density Analysis of Randomly Switched Pulse Width Modulation for DC/AC Converters

R. Lynn Kirlin\*, M. M. Bech\*\*, A M. Trzynadlowski\*\*\*

\*University of Victoria, Victoria, BC, Canada V8W 3P6, kirlin@ece.uvic.ca

\*\*University of Aalborg, DK 9220 Aalborg East, Denmark

\*\*\*University of Nevada Reno, Reno, NV 98557 USA

## I. Introduction

One of the new and inexpensive [1-6] concepts in power electronics is the principle of random pulse-width modulation (RPWM) for control of hard-switched power converters, accelerated by the steadily increasing concern with or regulations regarding emissions of acoustic noise, vibrations and electric fields. The random switching frequency (RSF-PWM) method has proven to be the most effective for reduction of acoustic annoyance, but due to the irregular sampling of samples in time, the method has not heretofore been accurately analyzed, and selection of the random switching frequencies has been more or less based on trial-and-error. This paper removes a great deal of the guesswork by providing formulas for not only the continuous spectrum (Watts/Hz), but also the pure power (Watts) components (harmonics) in both single-phase and three-phase voltage inverters, and these are verified by laboratory measurements.

## II. Principles of Random Switching Frequency PWM; from DC/DC to DC/AC

The power circuit of a three-phase inverter consists of three legs, which requires three independently controlled bilevel, single phase switching functions  $a(t)$ ,  $b(t)$ , and  $c(t)$ . The line-to-line voltage may be found as the difference between the switching functions e.g.  $u_{ab} = a(t) - b(t)$ . The time varying duty cycle of each pulse is  $M(t) = \frac{1}{2}(1 + mF(t))$ , where  $m$  is the modulation index and for sinusoidal modulation  $F(t) = \sin(2\pi f_1 t)$ , and for third harmonic injection

$F(t) = \frac{2}{\sqrt{3}}(\cos(2\pi f_1 t) - \frac{1}{6}\cos(6\pi f_1 t))$ . Each new pulse width is determined by a sample of  $F(t)$  taken at a time determined by a correspondingly new selection of the random switching frequency, as shown in figure 1. A summary of many randomization pulse schemes is given in [10], and included is a simple result of importance for DC/DC conversion which was

first analyzed in [8]. With considerable analytical effort we herein extend DC/DC results to DC/AC.

The fundamental process for DC/DC conversion is a random segment width (random switching frequency) pulse train, similar to that in figure 1, having sequential segments with random widths  $\tau_r$  and constant duty cycle  $0 \leq M \leq 1$  within the segments. When the segments start at times  $t_m$  instead of being centered as in figure 1, the process can be written

$$a(t) = \sum_{m=-\infty}^{\infty} u_m(t - t_m),$$

$$u_m(t) = \begin{cases} 1, & 0 \leq t \leq M\tau_m \\ 0, & \text{otherwise} \end{cases} \quad (1.1)$$

where the time location of the  $m^{\text{th}}$  pulse  $u_m(t - t_m)$  is

$$t_m = \sum_{r=0}^{m-1} \tau_r, \quad m = 1, 2, \dots, \quad \tau_0 = 0, \quad \text{where the}$$

segment widths  $\tau_r$  are randomly selected from a known distribution. When  $M$  is a deterministic periodic function of time, the process can be used for DC/AC conversion.

The power spectral density  $W(f)$  of a random process is defined to be the *time average*, as the time window duration  $T'$  approaches infinity, of the *ensemble average* of the magnitude squared of the Fourier transform of a time-windowed sample process [9]:

$$W(f) = \lim_{T' \rightarrow \infty} \frac{1}{T'} E\{|F[a_{T'}(t)]|^2\}, \quad (1.2)$$

where  $F[a_{T'}(t)]$  is the Fourier transform of the  $T'$ -second length of a sample process from the ensemble. Assuming independence of pulse widths in the  $m^{\text{th}}$  and  $(m+k)^{\text{th}}$  segment, Middleton [8] used this approach to write the power spectral density of  $a(t)$  produced the sum of the three spectral portions determined by the values of the correlation delay index  $k$  relative to the

sequence index  $m$ . With DC/AC conversion the duty cycle of the pulses depends upon the sampling point within the deterministic modulating waveform, and we now have

$$u_m(t) = \begin{cases} 1, & 0 \leq t \leq M(t_m)\tau_m \\ 0, & \text{otherwise} \end{cases} \quad (1.3)$$

showing that widths of any distinct pulses in the sequence are actually dependent, and Middleton's formula does not apply directly. We now consider that the duty cycles  $M(t)$  can be written as a periodic function of the uniform random phase variable  $\theta$ . The width of the  $m^{\text{th}}$  segment's pulse is now  $M(t_m)\tau_m$  a deterministic function. Then without loss of generality, because we have stationary, ergodic processes, we may as shown in figure 1 select  $t_m = 0$  as the sampling time  $t_m$  of the  $m^{\text{th}}$  segment and let the pulses be centered on, rather than start at, the sampling times.

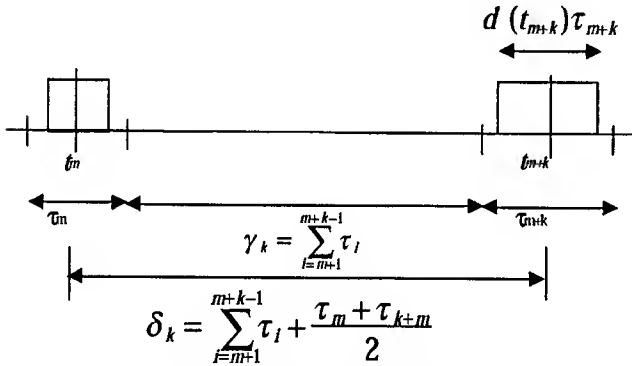


Figure 1. Example sequence of segments,  $k \geq 1$ .

Referring to figure 1, we can see that the time  $\gamma_k = t_{m+k} - t_m$  between the  $m^{\text{th}}$  and  $k^{\text{th}}$  pulse, if  $k \geq 1$ , is the sum of  $(k-1)$  random segment widths. The time  $\delta_k$  between pulse centers adds half the width of the  $m^{\text{th}}$  segment plus half the width of the  $m+k^{\text{th}}$  segment.

Thus

$$\gamma_k = \sum_{i=m+1}^{m+k-1} \tau_i, k \geq 2; \quad \gamma_k = 0, k = 1; \quad \gamma_k = -\tau_m, k = 0 \quad (1.4)$$

Finding the cross spectrum  $S_{m, m+k}$  between two pulses at times  $m$  and  $m+k$  over all  $k$  and taking expectation leads to the exact expression for the power density spectrum:

$$W(f) = \frac{\epsilon}{\bar{\tau}(\pi f)^2} E\{\sin^2(\pi f \tau_m M(t_m, \theta))\} + \frac{2\epsilon}{\bar{\tau}(\pi f)^2} \text{Re} \left\{ \sum_{k=1}^{\infty} E \left[ \sin \left( \pi f \tau_{m+k} M \left( \frac{t_m + \gamma_k}{2}, \theta \right) \right) e^{j\omega \frac{\tau_m + \tau_{m+k}}{2}} \right] \right\} \quad (1.5)$$

$\epsilon = 1, f = 0; \quad \epsilon = 2, f > 0$

### III. Approximation to the Exact formula for the Power Density Spectrum

The approximation by Bech [7],

$$\delta_k = \gamma_k + \frac{\tau_m + \tau_{m+k}}{2} \cong k\bar{\tau} \text{ in } M(\bullet) \text{ in the second}$$

factor of each term in the second expectation in (1.5), makes the corresponding factor independent from the other two and calculation possible. We implement the approximation and use the fact that  $M((k\bar{\tau} + LT), \theta) = M(k\bar{\tau}, \theta)$  and  $\gamma_K = \gamma_{LN+k}$ , where  $L = \text{fix}[K/N]$  and  $N = T/\bar{\tau}$ , and we also replace the resulting infinite geometric series sum over

$$L \text{ with } E_{\tau} \left\{ e^{j2\pi f \tau} \right\}^{k-1} \frac{1}{1 - E \left\{ e^{j\omega \tau} \right\}^{T/\bar{\tau}}}.$$
 Now

completely rewriting (1.5), we have the Bech approximation:

#### Single phase:

$$W(f) \cong \frac{\epsilon}{\bar{\tau}(\pi f)^2} E_{\theta, \tau_m} \left\{ \sin^2(\pi f \tau_m M(0, \theta)) \right\} + \frac{2\epsilon}{\bar{\tau}(\pi f)^2} \cdot \text{Re} \left\{ \sum_{k=1}^N E_{\theta} \left[ \begin{aligned} & E_{\tau_{m+k}} \left[ \sin(\pi f \tau_{m+k} M(0, \theta)) e^{j\omega \frac{\tau_m}{2}} \right] \\ & E_{\tau_{m+k}} \left[ \sin(\pi f \tau_{m+k} M(k\bar{\tau}, \theta)) e^{j\omega \frac{\tau_m + \tau_{m+k}}{2}} \right] \\ & E_{\tau} \left\{ e^{j\omega \tau} \right\}^{k-1} \end{aligned} \right] \right\} \left( 1 - E_{\tau} \left\{ e^{j\omega \tau} \right\}^N \right)^{-1} \quad (1.6)$$

### Line-to-line:

$$W_{b-a}(f) \equiv \frac{4}{\bar{\tau}(\pi f)^2} \left[ E_{\theta, \tau_m} \left\{ \begin{aligned} &\sin^2(\pi f \tau_m M(0, \theta)) \\ &-\sin(\pi f \tau_m M(0, \theta - 2\pi/3)) \sin(\pi f \tau_m M(0, \theta)) \end{aligned} \right\} \right. \\ \left. + 2 \operatorname{Re} \sum_{k=1}^N E_{\theta} \left\{ \begin{aligned} &E_{\tau_m} \left\{ \begin{aligned} &\sin(\pi f \tau_m M(0, \theta)) \\ &-\sin(\pi f \tau_m M(0, \theta - 2\pi/3)) \end{aligned} \right\} e^{j\omega \frac{\tau_m}{2}} \\ &\cdot E_{\tau_{m+k}} \left\{ \sin(\pi f \tau_{m+k} M(k\bar{\tau}, \theta)) e^{j\omega \frac{\tau_{m+k}}{2}} \right\} \\ &\cdot E_{\tau} \{ e^{j\omega \tau} \}^{k-1} \end{aligned} \right\} \right] \left( 1 - E_{\tau} \{ e^{j\omega \tau} \}^N \right)^{-1} \quad (1.7)$$

where  $\bar{\tau}$  is the expected segment width and  $N = T/\bar{\tau}$  (rounded).

## IV. Analysis of the Discrete Power Spectrum

Starting with a repeat of (1.5), retaining generality by setting sequence index  $m=0$  and letting  $t_m=0$ , treating  $M(\delta_k, \theta) = M(t_m + \gamma_k + \frac{\tau_m + \tau_{m+k}}{2}, \theta)$  as a deterministic function in the variable  $\delta_k = \gamma_k + \frac{\tau_m + \tau_{m+k}}{2}$  having period  $T = 1/f_1$ , replacing  $\sin(\pi f \tau_{m+k} M(\delta_k, \theta))$  with its exponential Fourier series having  $n^{\text{th}}$  coefficient  $F_n(f_d \tau_k) e^{jn\theta}$  and collecting the exponential delay factors gives

$$W(f) = \frac{\varepsilon}{(\pi f)^2 \bar{\tau}} \left[ E \left\{ \sin^2(\pi f \tau_0 M(0, \theta)) \right\} \right. \\ \left. + 2 \operatorname{Re} \left\{ \sum_{k=1}^{\infty} E \left\{ \begin{aligned} &\sin(\pi f \tau_0 M(0, \theta)) \\ &e^{j2\pi f \delta_k} \sum_{n=-\infty}^{\infty} \left( F_n(f_d \tau_k) e^{-j2\pi n f \delta_k} e^{-jn\theta} \right) \end{aligned} \right\} \right\} \right] \quad (1.8)$$

where  $\varepsilon = 1, f = 0; \varepsilon = 2, f > 0$ . Now we can see that the exponential factors  $e^{-j2\pi n f \delta_k}$  in the Fourier series and that of the delay factor  $e^{j2\pi f \delta_k}$  may combine for some discrete frequencies  $f = f_d$  to give unity for all  $k$ , causing the sum over  $k$  to give infinite power spectral density but finite power. Discrete

frequencies  $f = f_d$  having infinite power density have been shown to exist not only at 1) harmonics of  $f_{LCM}$ , the LCM of the set of switching frequencies and the factor 2, but also at 2) other discrete frequencies with non-zero power at  $f_1$ -spaced sidebands around harmonics of  $f_{LCM}$ . At these frequencies we use the expression for the power rather than power density. The resulting power expressions:

### single phase

$$P_{f_d} = \frac{2\varepsilon}{(\pi f \bar{\tau})^2} \left| E_{\tau} \{ F_n(f_d \tau) \} \right|^2 \quad (1.9)$$

### line-to-line

$$P_{f_d} = \frac{2\varepsilon}{(\pi f_d \bar{\tau})^2} \left\{ \left| E_{\tau_k} \{ F_{n_d}(f_d \tau_k) \} \right|^2 (1 - \cos(2n_d \pi / 3)) \right\},$$

$$\varepsilon = 1, f = 0; \varepsilon = 2, f > 0;$$

$$f_d = K f_{LCM} \pm n f_1; K, n = 0, 1, 2, \dots$$

(1.10)

where  $e^{jn\theta} F_n(f_d \tau)$  is the  $n^{\text{th}}$  Fourier coefficient in the exponential series expansion of  $\sin(\pi f \tau_k M(\delta_k, \theta))$  considering  $\delta_k$  the time variable with period  $T$ . We have also derived an analytic solution for the special case of modulation with a single sinusoid, where the  $F_n(f_d \tau)$  are easily found in terms of Bessel functions.

## V. Examples of Mixed Power and Power Density

Now that we have the expressions for approximate density, which is adequate at frequencies other than  $f_d = K f_{LCM} \pm n f_1$ , and for power that is accurate at discrete frequencies  $f_d = K f_{LCM} \pm n f_1$ , we may plot the two separately or in conjunction, mixing power density and power scales on the vertical axis. An example single phase approximate density computation of (1.6) is given in figure 2(a) where the fundamental frequency  $f_1 = 40$  Hz,  $m = 0.8$ , and the switching frequencies are 2.0, 2.5, 3.0, 3.5, and 4.0 kHz. Results are plotted with minimum frequency 500 Hz and spacing at 250 Hz. Because the fundamental, low harmonics and the  $f_{LCM}$  frequencies are out of range of the plot, there are no discrete harmonics to calculate here. This approximation is validated by the laboratory measurements shown in figure 2(b).

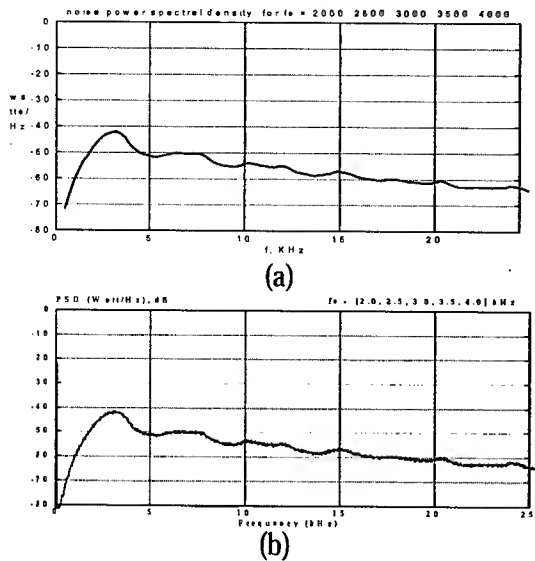


Figure 2 (a) Spectrum calculated from (1.6). Switching frequencies [2.0, 2.5, 3.0, 3.5, 4.0] kHz, having equally likely occurrence in time, and duty cycle modulation  $M(t) = \frac{1}{2}(1 + 0.8 \cos(\omega_1 t))$ . (b) Measurement of the same spectrum.

Figure 3 compares computations to measurements for line-to-line voltage near  $f_{LCM} = 12\text{KHz}$  for modulation by first and third harmonics of 40 Hz, switching frequencies at 2,3,4 kHz,  $m = 0.8$ . Figure 3(a) is the combination of 1) a dashed curve for the approximate density formula even wrongly calculating density at the discrete power frequencies, and 2) a solid curve giving the power formula at the discrete frequencies. The vertical scale is watts per hertz for the density, but watts for the discrete power frequencies. Figure 3(b) is generated from measurements with a *density* setting; it matches the approximate expression for the density except at those frequencies that relate to discrete harmonics. Figure 3(c) is generated from measurements with a power setting and matches exact power formula very well at the discrete frequencies. Both the density and power measurements fully verify the analysis. It should be emphasized that each part of these inherently mixed spectra require measurements with the proper PSD-scaled and a PWR-scaled settings.

## VI. Conclusions

This paper has presented a detailed spectral analysis of the random switching frequency PWM technique for control of three-phase inverters. A number of novel results have been reported including criteria for the existence of discrete harmonics in the output voltage waveforms, exact expressions for the power carried by these harmonics and an approximate but accurate

formula for the continuous density. Furthermore, the theory is generalized to any periodic modulation function. As an example, the third harmonic injection technique was analyzed, and based on extensive comparisons of predicted spectra with measured spectra and as opposed to all earlier investigations of random PWM schemes, the theory is found to be very accurate for both single phase and line-to-line spectra.

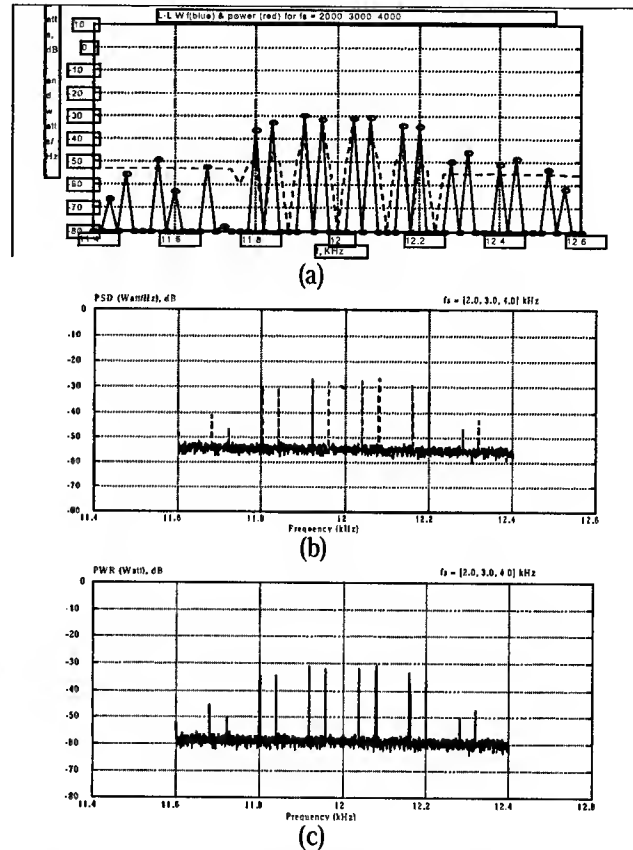


Figure 3. Comparisons near  $f_{LCM}$  for modulation by first and third harmonics of 40 Hz, line-to-line voltage, switching frequencies at 2,3,4 kHz,  $m = 0.8$ . (a) Power (circled, watts) and power spectral density (solid, watts/Hz); (b) measurement using density scaling; (c) measurement using power scaling.

## References

- [1] G. A. Covic and J. T. Boys, "Noise Quieting with Random PWM AC Drives," IEE Proc. Electric Power Applications, Vol. 145, No. 1, Jan 1998, pp. 1-10.
- [2] T. G. Habetler and D. M. Divan, "Acoustic Noise Reduction in Sinusoidal PWM Drives Using a Randomly Modulated Carrier," IEEE Trans. on Power Electronics, Vol. 6, No. 3, July 1991, pp. 356-363.
- [3] L. Xu, Z. Q. Zhu, and D. Howe, "Effect of Zero Space Vector and PWM Carrier on Acoustic Noise

from Induction Motor Drives," Proc. of the 32nd Universities Power Engineering Conference, Vol. 1, 1997, pp. 209-212.

[4] J. K. Pedersen and F. Blaabjerg, "Digital Quasi-Random Modulated SFAVM PWM in an AC-Drive System," IEEE Trans. on Industrial Electronics, Vol. 41, No. 5, Oct. 1995, pp. 518-525.

[5] M. M. Bech, F. Blaabjerg, and J. K. Pedersen, "Random Modulation Techniques with Fixed Switching Frequency for Three-Phase Power Converters", Accepted for presentation at the 1999 Power Electronics Specialists Conference in Charleston.

[6] F. Blaabjerg, J. K. Pedersen, L. Oestergaard, R. L. Kirlin, A. M. Trzynadlowski, and S. Legowski, "Optimized and Non-Optimized Random Modulation Techniques for VSI Drives," EPE Journal, Vol. 6, No. 2, 1996, pp. 46-53.

[7] Bech, M.M., 1999. Random Pulse-Width Modulation Techniques for Power Electronic Converters, Ph.D. Thesis, Aalborg University, Denmark.

[8] Middleton, D., 1987. Introduction to Statistical Communication Theory, Peninsula Publishing, Los Altos. (reprint from 1960, McGraw-Hill)

[9] Peebles, P.Z., 1993. Probability, Random Variables, and Random Signal Principles, 3<sup>rd</sup> Ed., McGraw-Hill, N.Y.

[10] Stankovic, A. M., 1993. Random Pulse Modulation with Applications to Power Electronic Converters, PhD Thesis, Massachusetts Institute of Technology, Feb.

# Study on Spectral Analysis and Design for DC/DC Conversion Using Random Switching Rate PWM

R. Lynn Kirlin, Jian Wang, and R. M. Dizaji

Department of Electrical and Computer Engineering, University of Victoria  
P.O.Box 3055, Victoria, B.C. V8W 3P6 Canada  
Email: jwang@ece.uvic.ca, kirlin@enr.uvic.ca

## Abstract

In this paper we give the formulation of the power spectral density of the randomized pulse width modulation (RPWM) DC/DC converter based on both constant duty cycle and constant pulse width schemes. We analyze the spectral formulas based on the constant duty cycle and develop a means for nulling the power spectral density at one specified frequency and its harmonics. We revert to optimization methods when switching frequency range is subject to practical constraints. Simulation results illustrate the effectiveness of our approach.

## 1. Introduction

The switching which is needed when converting a DC power source to a lower voltage typically causes both harmonics and electromagnetic emissions. It is possible to reduce or eliminate both of these by randomizing the switching, but it is may also be important to null the generated spectrum at frequencies for which the load has natural resonance. Proper design of the switching frequencies and their probability density function (PDF) gives control over the output noise power spectral density (PSD) at specified frequencies whose values we want to minimize. So the problem addressed by this research is the analysis and design of the power spectral density of randomized pulse width modulation (PWM) DC/DC converters by specification of an appropriate set of switching frequencies and their probabilities. Many randomization schemes for pulse width modulation (RPWM) in power converters have been proposed and several analyzed in detail in the literature. A summary of many of these, especially for DC/DC, is given by Stankovic [1]. Many schemes for DC/AC have also been published, but we are not concerned with those in this study. A simple result of importance for DC/DC conversion was given and analyzed first by Middleton [2] and also discussed by Stankovic [1]. Stankovic, Verghese, and Perrault [3] have addressed the DC/DC convertor problem of random switching design for spectral control of harmonic powers and cumulative power in specified bands, also our eventual intention. They used 2 switching

frequencies selected at random, but with Markov statistics such that long sequences of one switching frequency were discouraged. They designed for optimal results. They found the method ineffective for wideband control but successful for narrowband. Our approach is quite different, designing only a zero order switching frequency selection, and entirely working from the formulation in the frequency rather than the time (autocorrelation) domain. We consider specification of a minimum switching frequency to control worst case ripple in the low probability case of continual selection of the same switching frequency, and we consider specification of a maximum to control switching inefficiency and other circuit losses. In addition, we give new and powerful design results regarding the powers of the harmonics related to the distribution of the switching frequencies.

We first give the formulas of the two schemes, constant duty cycle (CDC) and constant pulse width (CPW) derived straightforwardly from the early work in [1]. Analysis of the formulas leads to a clear means of nulling the power spectral density at one specified frequency and its harmonics. At the same time the least common multiple frequency  $f_{LCM}$  of the switching frequencies can be designed as needed and the powers of  $f_{LCM}$  and its harmonics are easily determined.

Although formulation shows that the fixed pulse width scheme allows a spectral zero to be placed as desired within the constraints of the minimum, maximum and average switching frequency allowed, that scheme does not provide control of the duty cycle within each segment and is deemed "uncontrollable" because with some probability a long string of pulses with low or high duty cycles may momentarily cause unacceptable voltage deviations from the desired output DC level. Thus we focus our analysis and simulations on the formulas for the spectral density using the constant duty cycle scheme.

## 2. Formulations

We first present the summary of the equations for the power at discrete frequencies and the power spectral density (PSD) for both schemes CDC and CPW. General

derivations for the PSD are straightforward applications of formulas found in [1]. The derivations of the DC power and power at harmonics of the least common multiple  $f_{LCM}$  of the set of random switching frequencies have been derived in our reports to be published in more expanded form than allowed here. The  $f_{LCM}$  and its harmonics are those having finite power; this can be seen by noting which frequencies cause the denominator terms  $1 - E(e^{j2\pi f \tau})$  in the PSDs of (2) and (4) below to be zero, assuming a finite set of switching frequencies are randomized.

## 2.1. Formulas for CPW Scheme

For CPW every pulse has constant amplitude  $A_0$  and duration  $\tau_0$ , but the average switching interval is such that the average duty cycle  $\alpha = \tau_0 / \bar{\tau}$  is as desired.  $E$  denotes statistical average, and  $\tau$  is the random segment duration, the reciprocal of the random switching frequency, and  $E\{\tau\} = \bar{\tau}$ . The frequencies  $f^*$  denote the  $f_{LCM}$  or any one of its harmonics.

Power at discrete frequencies (watts)

$$P(f^*) = \begin{cases} P_{DC} = \left( \frac{A_0 \tau_0}{\bar{\tau}} \right)^2, & f^* = 0 \\ \frac{2A_0^2}{(\bar{\tau} \pi f^*)^2} \sin^2(\pi f^* \tau_0), & f^* > 0 \end{cases} \quad (1)$$

One-sided spectral density (watts/Hz)

$$W_{ys}(f) = \frac{\varepsilon (A_0 \sin(\pi f \tau_0))^2}{\bar{\tau} \pi^2 f^2} \left( 1 + 2 \operatorname{Re} \left[ \frac{E(e^{j2\pi f \tau})}{1 - E(e^{j2\pi f \tau})} \right] \right), \quad \varepsilon = 1, f = 0; \varepsilon = 2, f > 0 \quad (2)$$

## 2.2. Formulas for the CDC Scheme

In this scheme the pulses have varying width, according to the width of the segment but the duty cycle is fixed. The same notation is used as in CPW.

Power at discrete frequencies (watts)

$$P(f^*) = \begin{cases} (A_0 \alpha)^2, & f^* = 0 \\ S_y(f^*) = \frac{2A_0^2}{(\pi f^* \bar{\tau})^2} \left| E\{\sin(\pi \alpha f^* \tau) e^{-j\pi \alpha f^* \tau}\} \right|^2 \\ = \frac{2A_0^2}{(2\pi f^* \bar{\tau})^2} \left| E\{1 - e^{-j2\pi \alpha f^* \tau}\} \right|^2, & f^* > 0 \end{cases} \quad (3)$$

One-sided spectral density (watts/Hz)

$$W_{ys}(f) = \frac{\varepsilon A_0^2}{\bar{\tau} \pi^2 f^2} \left( \frac{E\{\sin^2(\pi \alpha f \tau)\}}{+ 2 \operatorname{Re} \left[ \frac{E\{\sin(\pi \alpha f \tau) e^{j\pi f \tau}\}}{1 - E\{e^{j2\pi f \tau}\}} \right]} \right), \quad \varepsilon = 1, f = 0; \varepsilon = 2, f > 0 \quad (4)$$

## 3. Spectral Analysis and Design Techniques for CDC Scheme

Now we must choose the proper switching frequencies and their probabilities for control of the discrete frequencies having non-zero power and for minimization of the power spectrum at a desired frequency  $f_0$ . Both design goals are oriented toward keeping generated power away from frequencies known to be harmful either to the environment or converter loads. It is assumed that the designs can be updated adaptively.

### 3.1. Nulling $E\{\sin^2(\pi \alpha f \tau)\}$ and $E\{\sin(\pi \alpha f \tau) e^{j\pi f \tau}\}$

Because a null is repeated periodically on the frequency axis, we direct our attention to the lowest possible frequency that can be nulled given the constraints on the range of switching frequencies. Both first and second terms in (4) are nulled if for the given  $\alpha$  the random variables  $\tau_i$  are selected such that the values of  $\sin^2(\pi \alpha f \tau)$  and  $\sin(\pi \alpha f \tau) e^{j\pi f \tau}$  are always zero at  $f = f_0$ . Thus writing  $f_0 = \beta \bar{f}_s$ , where  $\bar{f}_s$  is the average switching frequency and  $\beta$  is a positive scale factor, we can see that for a null at  $f_0$  (and all its

harmonics,  $k = 1, 2, \dots$ ) the relationship that the switching frequencies  $f_{s_i} = 1/\tau_i$  must have to  $\bar{f}_s$  is:

$$f_{s_i} = \frac{\alpha \beta \bar{f}_s}{k} = \frac{\alpha f_0}{k}, i = 1, 2, \dots, N_s; k = 1, 2, \dots \quad (5)$$

assuming  $N_s$  is the maximum number of switching frequencies used. Because the  $f_{s_i}$  are also constrained to lie in  $[f_{s_{\min}}, f_{s_{\max}}]$  and the largest must be greater than the average, the largest  $f_{s_i} = f_{s_1}$  in (5) is constrained by

$$\bar{f}_s \leq f_{s_1} = \alpha \beta \bar{f}_s / k_1 \leq f_{s_{\max}} \quad (5.1)$$

where  $k_1$  gives the relationship (5).

In order for the average to be as desired,  $f_{s_1}$  must be given a probability that makes its weight greater than or equal to that of all the other frequencies. Similarly, the smallest  $f_{s_i} = f_{s_{N_s}}$  is constrained by

$$f_{s_{\min}} \leq f_{s_{N_s}} = \alpha \beta \bar{f}_s / k_{N_s} \leq \bar{f}_s \quad (5.2)$$

### 3.1.1. Limitations on $f_0$ , the Null Frequency for $E\{\sin^2(\pi \alpha f \tau)\}$ and $E\{\sin(\pi \alpha f \tau)e^{j\pi f \tau}\}$

If the constraints (5) – (5.2) cannot be met, then no design can give a pure null in both terms of (4). In fact, (5.1) and (5.2) jointly give limits on the lowest frequency of null  $f_0 = \beta \bar{f}_s$ :

$$f_{s_{\min}} \leq f_{s_{N_s}} = \alpha f_0 / k_{N_s} \leq \bar{f}_s \leq f_{s_1} = \alpha f_0 / k_1 \leq f_{s_{\max}} \quad (6)$$

Placing nulls at higher frequencies is not as difficult, because the frequencies of nulls are periodic, and we may null any frequency  $k f_0$  as long as  $f_0$  satisfies the above expressions. The lowest switching frequency can easily satisfy (6) if  $f_{s_{\min}}$  is not too close to  $f_{s_{\max}}$ .

### 3.1.2. Simulation Results and an exception

We first choose a set of switching frequencies according to (5), nulling  $E\{\sin^2(\pi \alpha f \tau)\}$  with a duty cycle of 80%, switching frequencies 25KHz and 50KHz,  $f_0 = 62.5\text{KHz}$ . The resulting PSD value at  $f_0$  is -140.1 dB. The spectral result is shown in figure 1.

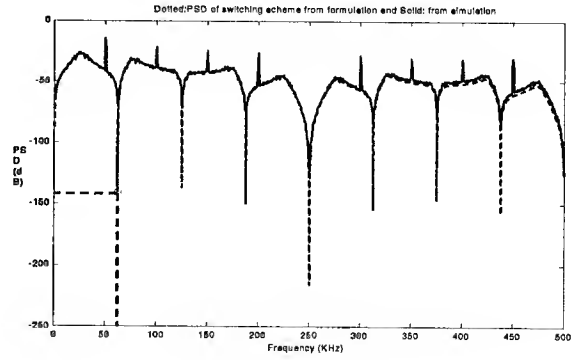


Figure 1. Results for duty cycle 80%, switching frequencies 25KHz and 50KHz. Dotted: theoretical. Solid: simulation. There is ideally a perfect null at 62.5KHz.

If we choose duty cycle equal to 50% and switching frequencies 15.6KHz and 31.3KHz, and  $f_0 = 62.5\text{KHz}$  then we get the result shown in figure 2. Note that the theoretical PSD value at  $f_0 = 62.5\text{KHz}$  from the formula is -172.6dB, but the simulation's value at this frequency is about -45dB. This difference is due to a non-determinism

that will always occur when duty cycle  $\alpha = \frac{1}{2}, \frac{1}{3}, \dots, \frac{1}{k}$ .

Thus for this coincidental case we cannot use the nulling of  $E\{\sin^2(\pi f \alpha \tau)\}$  to get a sharp null at the desired frequency  $f_0$  if  $f_0$  coincides with a harmonic of the least common multiple  $f_{LCM}$ .

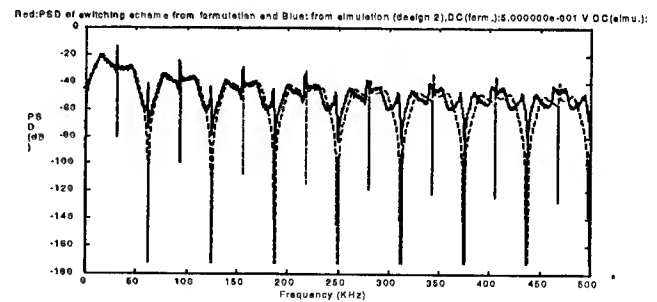


Figure 2. Nulling  $E\{\sin^2(\pi \alpha f \tau)\}$  for 50% duty cycle, switching frequencies 15.6KHz and 31.3KHz,  $f_0 = 62.5\text{KHz}$ . Dotted: analytical. Solid: simulation.

## 3.2. Maximizing $1 - E\{e^{j2\pi f \tau}\}$



The magnitude of the second term in (4) at frequencies other than those nulled by (6) is dominated by the denominator expression,  $1 - E\{e^{j2\pi f\tau}\}$ , which can approach zero and cause the density not only to become large but perhaps to become infinite. To preclude that, we can force  $1 - E\{e^{j2\pi f\tau}\}$  to have maximum magnitude at a desired frequency  $f_0' = \beta \bar{f}_s$  as above, except that we are denoting  $f_0'$  as the frequency at which we would like to maximize  $1 - E\{e^{j2\pi f\tau}\}$ . Every exponential term can precisely equal -1 if  $2\pi f_0' / f_{s_i} = (2k+1)\pi$ ,  $k = 0, 1, 2, \dots, N_s$ . Thus we have that:

$$f_{s_k} = \frac{2\beta \bar{f}_s}{2k-1} = \frac{2f_0'}{2k-1}, \quad k = 1, 2, \dots, N_s \quad (7)$$

As before the minimum and maximum frequencies are constrained according to

$$\bar{f}_s \leq f_{s_1} = 2\beta \bar{f}_s \leq f_{s_{\max}} \quad (7.1)$$

$$f_{s_{\min}} \leq f_{s_{N_s}} = 2\beta \bar{f}_s / (2N_s - 1) \leq \bar{f}_s \quad (7.2)$$

### 3.2.1. Limitations on $f_0'$ , the Maximizing Frequency for $1 - E\{e^{j2\pi f\tau}\}$

The constraints imposed by minimum, maximum and average switching frequencies on the frequency of maximum of the denominator factor  $1 - E\{e^{j2\pi f\tau}\}$  are similar as those imposed by the term  $E\{\sin^2(\pi\alpha f\tau)\}$ .

$$\frac{f_{s_{\min}}}{2} \leq \frac{f_{s_{N_s}}}{2} = \frac{f_0'}{(2N_s - 1)} \leq \frac{\bar{f}_s}{2} \leq \frac{f_{s_1}}{2} = f_0' \leq \frac{f_{s_{\max}}}{2} \quad (8)$$

### 3.2.2. Contradiction: Simultaneously nulling $E\{\sin^2(\pi\alpha f\tau)\}$ and $E\{\sin(\pi\alpha f\tau)e^{j\pi f\tau}\}$ while maximizing $1 - E\{e^{j2\pi f\tau}\}$

If we combine the inequalities of the limitations it is clear that both objectives cannot be met simultaneously because

$$\text{for } 0 \leq \alpha \leq 1 \text{ we cannot have } \frac{f_{s_1}}{\alpha} = f_0' \text{ and } \frac{f_{s_1}}{2} = f_0'$$

equal to each other. Clearly an overall or global optimal must be found by functional minimization to constraints.

When we want to minimize the PSD within the constraints of the average and finite range of switching frequencies, the most important factor is  $1 - E\{e^{j2\pi f\tau}\}$ . This is because when we want to null  $\sin(\pi\alpha f\tau)$  at an arbitrary frequency there are practical limitations. We have also found that even without the constraints on switching frequencies sometimes  $\sin(\pi\alpha f\tau)$  cannot be simply nulled if the frequency of desired null coincides with a harmonic of the least common multiple  $f_{LCM}$  (figure 2).

## 3.2.3 Simulation Results and Discussion

When we maximize  $1 - E\{e^{j2\pi f\tau}\}$  to minimize the peak at  $f_0 = 62.5\text{KHz}$  according to (7), the switching frequencies for maximization can be 25KHz and 41.7KHz. The power spectral density is shown in figure 3.

## 4. Optimization Method and Results

We now apply nonlinear optimization method to find a constrained set of switching frequencies and their corresponding probabilities that will minimize the spectral power at a specified frequency. We formulate the minimization problem, apply optimization method and then give experiment results.

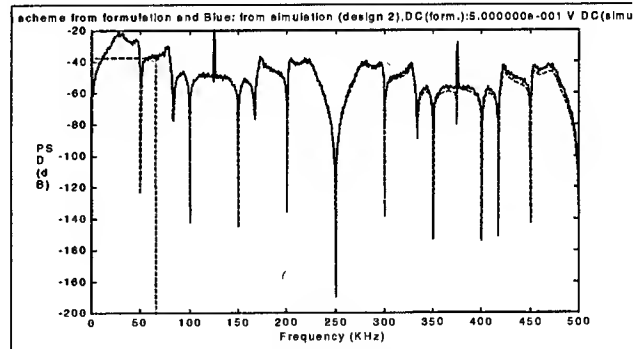


Figure 3. Maximizing  $1 - E\{e^{j2\pi f\tau}\}$  to minimize the peak at  $f_0 = 62.5\text{KHz}$  according to (7); switching frequencies are 25KHz and 41.7KHz. Dotted: analytical. Solid: simulation.

We can formulate our minimization problem from (4) as follows:

$$W_{ys}(f) =$$

$$\frac{2A_0^2}{\sum_{i=1}^m p_i \tau_i \pi^2 f_0^2} \left[ \begin{aligned} & \left( \sum_{i=1}^m p_i \left\{ \sin(\pi \alpha_0 f_0 \tau_i) \cos(\pi f_0 \tau_i) \right\} - \sum_{i=1}^m p_i \left\{ \sin(\pi \alpha_0 f_0 \tau_i) \sin(\pi f_0 \tau_i) \right\} \right) \\ & \times \left( 1 - \sum_{i=1}^m p_i \left\{ \cos(2\pi f_0 \tau_i) \right\} \right) \\ & - 2 \left( \sum_{i=1}^m p_i \left\{ \sin(\pi \alpha_0 f_0 \tau_i) \cos(\pi f_0 \tau_i) \right\} \right) \\ & \left( \sum_{i=1}^m p_i \left\{ \sin(\pi \alpha_0 f_0 \tau_i) \sin(\pi f_0 \tau_i) \right\} \right) \\ & \times \left( \sum_{i=1}^m p_i \left\{ \sin(2\pi f_0 \tau_i) \right\} \right) \\ & / \left[ \left( 1 - \sum_{i=1}^m p_i \cos(2\pi f_0 \tau_i) \right)^2 + \left( \sum_{i=1}^m p_i \sin(2\pi f_0 \tau_i) \right)^2 \right] \\ & + \frac{A_0^2}{\sum_{i=1}^m p_i \tau_i \pi^2 f_0^2} \sum_{i=1}^m p_i \sin^2(\pi \alpha_0 f_0 \tau_i) \end{aligned} \right] \quad (9)$$

Subject to the constraints:

$$\begin{cases} \frac{1}{75} \leq \tau_1, \tau_2 \dots \tau_m \leq \frac{1}{55} \\ p_1, p_2 \dots p_m \geq 0 \\ \sum_{i=1}^m p_i = 1, \sum_{i=1}^m \frac{p_i}{\tau_i} = 62.5 \end{cases} \quad (10)$$

The corresponding original and optimized power spectral densities for 5 random frequencies are shown together for comparison in figure 4. Note that the value at 62.5KHz has dropped about 3dB.

## 5. Conclusions

We have analyzed the power spectral formulas and give the ways to control the power at the desired frequency.

The simulation results prove the effectiveness of the concept and method. When we have some constraints on the switching frequencies, we can still able to minimize power e at some chosen frequency, although we may not be able to null it. Our future work will focus on control of the power spectral over a *band* of frequencies.

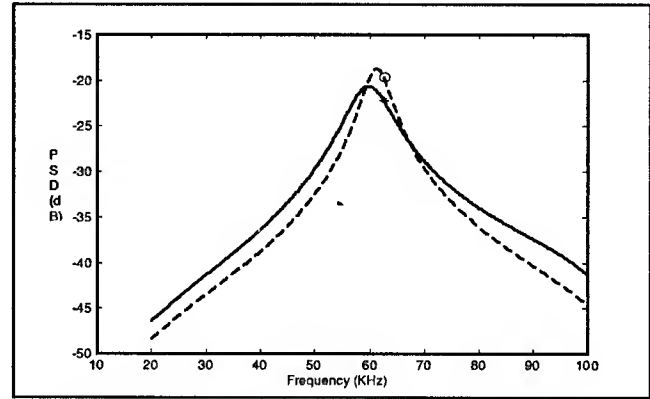


Figure 4. Power spectral density detail around  $f = 62.5\text{KHz}$ . Dotted line represents a random set of 5 switching frequencies. Solid line represents the optimized solution for those frequencies

## Acknowledgement

Part of this work was done under a contract from The Advanced Science Institute of British Columbia and Xantrex

## References

- [1] Stankovic, A. M., 1993. Random Pulse Modulation with Applications to Power Electronic Converters, PhD Thesis, Massachusetts Institute of Technology, Feb.
- [2] Middleton, D., 1987. Introduction to Statistical Communication Theory, Peninsula Publishing, Los Altos. (reprint from 1960, McGraw-Hill)
- [3] Stankovic, A. M., G.C. Verghese, and D.J. Perrault, Analysis and Synthesis of Random Modulation Schemes for Power Converters, Proc. 24<sup>th</sup> Ann. IEEE Power Electronics Specialists Conference, Seattle, 1993; IEEE cat. no. 93ch-3293-8, ISBN 0-7803-1243-0.

# SPECTRAL SUBTRACTION AND SPECTRAL ESTIMATION

Miguel A. Lagunas, Ana I. Perez-Neira

TSC Department, Modulo D5, Campus Nord UPC  
Jordi Girona 1-3, 08034 BARCELONA, SPAIN

## ABSTRACT

The problem of spectral subtraction, to estimate the parameters of a single source in colored noise, is used to show the relationships between the likelihood formulation and spectral density estimation. Reported previously as a filter bank processing for spectral estimation, it is shown that the normalized Capon estimate is the natural tool for source location in 1-d and 2-d scenarios when the noise background estimate is faced as a spectral subtraction problem. Several simulations selecting 1-d and 2-d apertures are used to show the degree of quality achieved with the proposed formulation. Also, the Periodogram test for incoherent detection is analyzed in front of the optimum test and the herein referred to as the Capon's test.

## I. INTRODUCTION

Using the maximum likelihood formulation for the problem of a line source embedded in colored noise, this work seeks the relationships between frequency detectors, spectral density estimates and spectral subtraction.

It seems clear that characterizing the source location, its power level and the spectral density of the noise, entails to estimate first the source location, second its power level and finally, by spectral subtraction, the noise spectral density. Assuming this path in the procedure, it seems also clear that a high resolution line detector is needed at the first step and this is the reason for the interest towards the spectral density estimates. For the second step, a power level estimate is required and apparently the Capon estimate has no competitor to perform such estimation. Finally, the third step reduces, again apparently, to the subtracting the estimated line contribution from the data covariance matrix. Regardless this protocol is valid in essence, there are several issues of interest, related with the maximum likelihood formulation, that preclude and arbitrary choice of the procedures used at every step.

To go through the mentioned steps, the problem of detecting a line source in colored noise has been selected. The presentation focuses on the case of an ULA array, leaving the case of 2-D apertures to the simulations section.

This work has been partially supported by the European Commission under Project IST-1999-11729 METRA; the Spanish Government (CYCIT) TIC99-0849, and the Catalan Government (CIRIT) 1998SGR 00081, 1999FI 00588.

The snapshot model we are assuming is formed by a single line impinging on a ULA array of  $Q$  sensors with colored noise  $\underline{w}(n)$ . It is considered that the actual complex envelope and the spatial frequency are  $\alpha_e(n)$  and  $f_0$  respectively, i.e.  $f_0$  is equal to the product of the inter-element distance in wavelength by the sinus of the source elevation  $d \cdot \sin(\theta_0)$ . The snapshot model is given by (1), where  $\underline{S}_e$  is the  $(Q \times 1)$  steering vector at frequency  $f_0$ .

$$\underline{X}_n = \alpha_e(n) \cdot \underline{S}_e + \underline{w}_n \quad (1)$$

The log likelihood of this data snapshot is given by (2), where  $\hat{a}_n$ ,  $\hat{\underline{S}}$  and  $\hat{\underline{R}}_o$  are the estimates of the complex envelope, the steering vector and the noise covariance matrix respectively.

$$\Lambda_n = L_n(\det[\hat{\underline{R}}_o^{-1}] - (\underline{X}_n - \hat{a}_n \hat{\underline{S}})^H \hat{\underline{R}}_o^{-1} (\underline{X}_n - \hat{a}_n \hat{\underline{S}})) \quad (2)$$

From this expression is obvious that the power level estimate is the suitable step to proceed first, instead the source location since the log-likelihood is highly non-linear on this parameter. The ML estimate of the source complex envelope is derived from the maximization of the above expression. The estimate is given by (3.a) and in (3.b) the corresponding power level estimate, where  $N$  is the number of collected snapshots.

$$\hat{a}_n = \frac{\underline{S}_e^H \hat{\underline{R}}_o^{-1} \underline{X}_n}{\underline{S}_e^H \hat{\underline{R}}_o^{-1} \underline{S}_e} \quad (3.a)$$

$$\hat{\alpha} = \frac{1}{N} \sum_{n=0}^{N-1} |\hat{a}_n|^2 = \frac{\underline{S}_e^H \hat{\underline{R}}_o^{-1} \hat{\underline{R}} \hat{\underline{R}}_o^{-1} \underline{S}_e}{(\underline{S}_e^H \hat{\underline{R}}_o^{-1} \underline{S}_e)^2} \quad (3.b)$$

Note that regarding to (3.b) that it looks different from the traditional Capon estimate [1]. Usually it is argued that the product of the inverse of the noise covariance matrix by the steering is proportional to the vector resulting of using the data matrix instead of the noise matrix. Nevertheless, this property depends on the estimates of the noise covariance matrix and the steering of the source, which at this step are not available.

## II THE LOG-LIKELIHOOD OF THE STEERING AND THE NOISE COVARIANCE

Using (3.a) in (2) and summing up for the available snapshots, the log-likelihood to be maximized is obtained.

$$\Lambda = \ln(\det(\underline{R}_{=o}^{-1})) - \text{trace} \left\{ \underline{R}_{=o}^{-1} \left[ \underline{R} - \frac{\underline{S} \underline{S}^H \underline{R}_{=o}^{-1} \underline{R}}{\rho_o} \right] \right\} \quad (4)$$

where

$$\rho_o = \underline{S}^H \underline{R}_{=o}^{-1} \underline{S} \quad (5)$$

Re-arranging terms in (4), it can also be written as (6).

$$\Lambda = \ln(\det(\underline{R}_{=o}^{-1})) - \text{trace} \left[ \underline{R}_{=o}^{-1} \underline{R} \right] + \frac{\underline{S}^H \underline{R}_{=o}^{-1} \underline{R} \underline{R}_{=o}^{-1} \underline{S}}{\underline{S}^H \underline{R}_{=o}^{-1} \underline{S}} \quad (6)$$

This last formulation reveals that the optimum detection test  $T_{opt}(f)$  to estimate the source location, assuming that the noise covariance is known, is a Raleigh quotient, which is close to the power estimate (3.b) derived previously. This test plays always an important role in improving detectors and beamformers [4]. It is also coherent with the white noise case since it coincides precisely with the data Periodogram.

Before going on with the procedure, it is worth studying the properties of the optimum test under perfect knowledge of the inverse noise covariance.

Assuming that the inverse of the noise covariance is known, the optimum test to locate the source steering is to maximize (7).

$$T_{opt}(f) = \frac{\underline{S}^H \underline{R}_{=o}^{-1} \underline{R} \underline{R}_{=o}^{-1} \underline{S}}{\underline{S}^H \underline{R}_{=o}^{-1} \underline{S}} \quad (7)$$

This test is lower bounded by using the following definitions an inequality:

$$\underline{v} = \underline{R}^{1/2} \underline{R}_{=o}^{-1} \underline{S} \quad ; \quad \underline{u} = \underline{R}^{-1/2} \underline{S} \quad ; \quad \|\underline{v}\|^2 \leq \frac{|\underline{u}^H \underline{v}|^2}{\|\underline{u}\|^2} \quad (8)$$

In consequence,

$$T_{opt}(f) = \frac{\underline{S}^H \underline{R}_{=o}^{-1} \underline{R} \underline{R}_{=o}^{-1} \underline{S}}{\underline{S}^H \underline{R}_{=o}^{-1} \underline{S}} \leq \frac{\underline{S}^H \underline{R}_{=o}^{-1} \underline{S}}{\underline{S}^H \underline{R}_{=o}^{-1} \underline{S}} = T_{Cap}(f) \quad (9)$$

This reveals that, in terms of resolution the right term of (9) will be better than the optimum test, since both get the same value at the true steering. Also it is very important to remark that the so-called classic test formed by the quotient of the periodograms measures do not bound, in any way, the optimum test. Its use is only suitable in the case of white noise only; in this case the Periodogram test coincides with the optimum test, which still is bounded by the Capon test.

$$T_{Period} = \frac{\underline{S}^H \underline{R} \underline{S}}{\underline{S}^H \underline{R}_{=o} \underline{S}} \quad (10)$$

Just to put in evidence the above comments, in Figure 1, they are represented simultaneously all the tests described under the condition of perfect knowledge of the noise covariance matrix. The data are formed by a source located at the spatial frequency 0.1, the colored noise obtained from a Moving Average MA(3) model, with model coefficients (1 0 -1), and SNR equal to 8 dB. This figure reveals the claimed superiority of the Capon test over the test derived from the log-likelihood. More important is the poor performance of the Periodogram test.

### III. SPECTRAL SUBTRACTION

An alternative to the direct maximization of the log-likelihood over the noise and steering parameters is to set the relationship between the noise covariance matrix and the steering to be estimated as a spectral subtraction problem. Note that noise covariance estimation is a major issue for power level estimation [2], and spectral subtraction, regardless of being heuristically in many cases, uses to be the suitable tool to perform it.

Under a spectral subtraction approach the noise covariance matrix is formulated as the subtraction of the data covariance matrix minus the source contribution .

$$\underline{R}_{=o,1} = \underline{R} - \beta \underline{S} \underline{S}^H \quad (11)$$

The maximum value of  $\beta$ , in order to ensure that the estimated matrix is positive definite, is precisely the traditional Capon estimate. The problem is that using this estimate precludes the use of the inverse since it does not exist because the minimum eigenvalue of the estimated noise matrix is zero.

$$\beta_{max} = \frac{1}{\underline{S}^H \underline{R}_{=o}^{-1} \underline{S}} \quad (12)$$

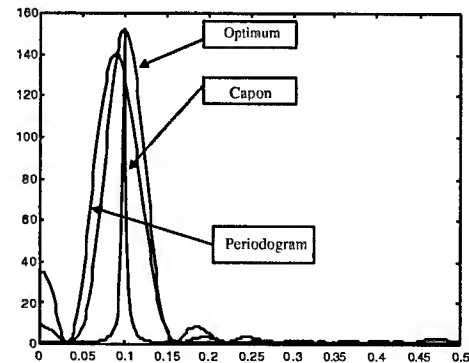


Figure 1. Optimum (—), Capon (---) and Periodogram test (···).

The second attempt is more suitable and faces directly the estimate of the inverse, since it is the only one that is required in the formulation of the log-likelihood when using the

appropriate test for source location. This estimate is given in (13), where the second term is the rank one contribution of the vector (norm one) that nulls the contribution of the last two terms of (6). Again, the parameter  $\gamma$  has to be bounded in order to preserve the positive character of the estimate.

$$\underline{\underline{R}}_o^{-1} = \underline{\underline{R}}^{-1} + \gamma \cdot \left( \frac{\underline{\underline{R}}^{-1} \cdot \underline{\underline{S}} \cdot \underline{\underline{S}}^H \cdot \underline{\underline{R}}^{-1}}{\underline{\underline{S}}^H \cdot \underline{\underline{R}}^{-2} \cdot \underline{\underline{S}}} \right) \quad (13)$$

This bound is given in (14), where it is clear that it is less restrictive than the case of modeling directly the noise matrix.

$$\gamma \geq - \frac{1}{\underline{\underline{S}}^H \cdot \underline{\underline{R}}^{-1} \cdot \underline{\underline{S}}} \quad (14)$$

At the same time, the log-likelihood for this noise estimate is equal to (15), in consequence  $\gamma$  has to be selected in order to maximize the determinant of the inverse noise covariance matrix.

$$\Lambda = \text{Ln}(\det(\underline{\underline{R}}_o^{-1})) - (Q-1) ; \forall \gamma \quad (15)$$

Finally, with this choice for the parameter, the log-likelihood, the optimum test, described in the previous section, and the so-called normalized Capon estimate [3] are easily related as follows:

$$\begin{aligned} \Lambda + Q - 1 &= \text{Ln}(\mathcal{T}_{opt}(f)) = \\ &= \text{Ln} \left[ \frac{\underline{\underline{S}}^H \cdot \underline{\underline{R}}_o^{-1} \cdot \underline{\underline{R}} \cdot \underline{\underline{R}}^{-1} \cdot \underline{\underline{S}}}{\underline{\underline{S}}^H \cdot \underline{\underline{R}}_o^{-1} \cdot \underline{\underline{S}}} \right] = \text{Ln} \left[ 1 + \gamma \frac{\underline{\underline{S}}^H \cdot \underline{\underline{R}}^{-1} \cdot \underline{\underline{S}}}{\underline{\underline{S}}^H \cdot \underline{\underline{R}}^{-2} \cdot \underline{\underline{S}}} \right] \end{aligned} \quad (16)$$

In summary, viewing the problem of the noise covariance matrix estimation as a problem of spectral subtraction, carried over the inverse of the data covariance matrix, reveals that the optimum frequency detector is not the classical minimum variance beamformer. It is the so-called normalized Capon spectral estimate which provides the optimum test for frequency detection.

It may be argued that in the above formulation the parameter  $\gamma$  may also depend on the steering vector, in this respect, next section will describe some empirical support to the choice of this parameter, setting it to a constant value. Note that this is equivalent to set a constant value for the trace of the inverse noise covariance matrix estimate, independently of the steering selected.

Before closing this section, in Figure 2 they can be viewed: Right, the likelihood using (16); and, left, the noise spectral estimate using (13) and the Capon power level estimate.

#### IV SPECTRAL ESTIMATION.

Rewriting again the noise covariance estimate leaving unknown the parameter  $K_o(\underline{\underline{S}})$ ,

$$\underline{\underline{R}}_o^{-1} = \underline{\underline{R}}^{-1} + K_o(\underline{\underline{S}}) \cdot \underline{\underline{R}}^{-1} \cdot \underline{\underline{S}} \cdot \underline{\underline{S}}^H \cdot \underline{\underline{R}}^{-1} \quad (17)$$

and assuming that the source is at this steering with power level  $\alpha_s$ , the correct value of  $K_o$  is (18), where  $\rho$  is the inverse of the power level from the Capon estimate.

$$K_o(\underline{\underline{S}}) = \frac{\alpha_s}{1 - \rho(\underline{\underline{S}})\alpha_s} \quad \text{with} \quad \rho = \underline{\underline{S}}^H \cdot \underline{\underline{R}}^{-1} \cdot \underline{\underline{S}} \quad (18)$$

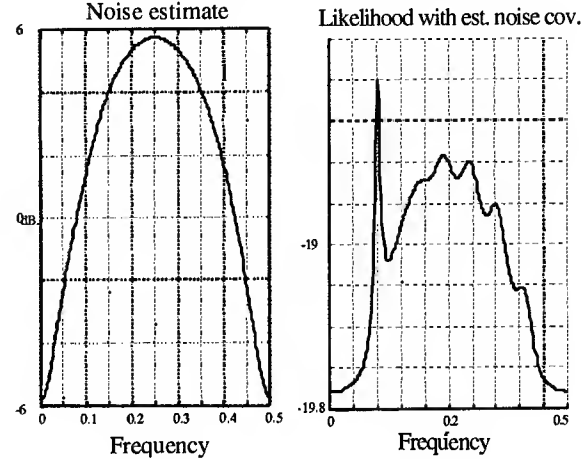


Figure 2. (Left). Capon Noise spectral density estimate. (Right). Log-Likelihood

Since the power level  $1/\rho$  contains both the signal and the noise contributions and, in addition, we assume that the noise power can be formulated by a white density  $N_o$  and a shaping bandwidth  $B(\underline{\underline{S}})$ , then

$$\frac{1}{\rho} = \alpha_s + \alpha_{on} = \alpha_s + N_o \cdot B(\underline{\underline{S}}) \quad (19)$$

In consequence, the product  $K_o(\underline{\underline{S}}) \cdot \rho$  can be formulated as (20), where  $\Psi(\underline{\underline{S}})$  is the spectral density.

$$K_o(\underline{\underline{S}}) \cdot \rho = \frac{1}{N_o} \cdot \frac{\alpha_s}{B(\underline{\underline{S}})} = \frac{1}{N_o} \cdot \Psi(\underline{\underline{S}}) \quad (20)$$

Furthermore, using (20) in the corresponding log-likelihood, results in (21), which proves the efficiency and the relationship between line detectors or spectral density estimates and the maximization of the log-likelihood. At the same time, taking into account that the normalized estimate provides spectral density, ensures that the proper choice for parameter  $\gamma$  in the previous section is a constant independent of the steering.

$$\Lambda - Q + 1 = \text{Ln}[1 + K_o(\underline{\underline{S}}) \cdot \rho] = \text{Ln} \left[ 1 + \frac{\Psi(\underline{\underline{S}})}{N_o} \right] \quad (21)$$

#### V. SIMULATIONS

In order to show that the framework described previously is also valid for 2-D problems, the hexagonal aperture depicted in Figure 3 has been selected for this section.

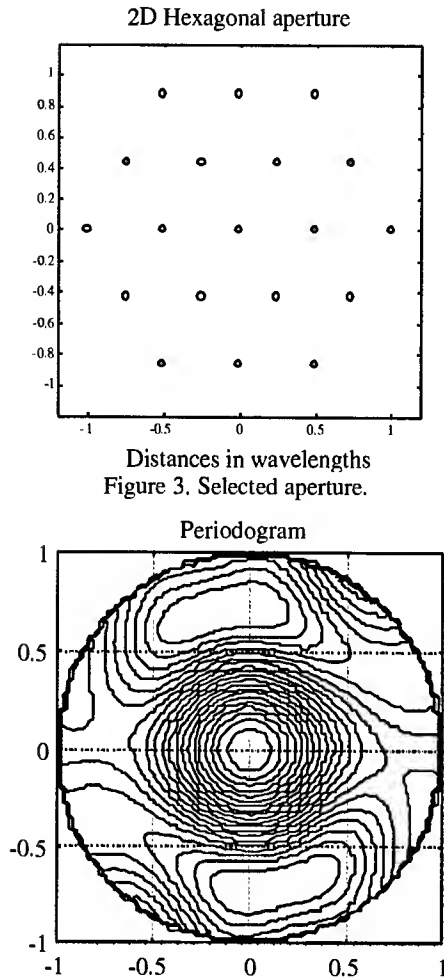


Figure 4. The noise Periodogram. Polar slowness-azimuth plot.

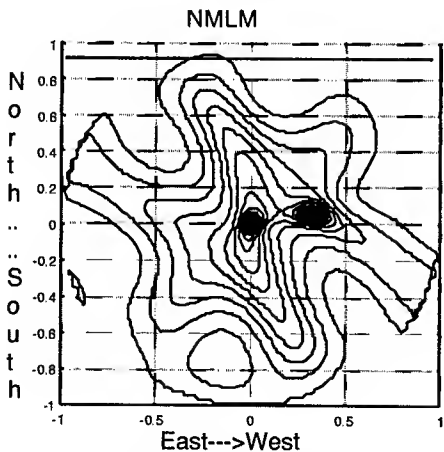


Figure 5. The Normalized estimate, proportional to the log-likelihood, for the scenario defined in the text.

The noise was spatially colored being its Periodogram estimate as depicted in Figure 4. The representation uses the slowness-azimuth plot being the south to north axis coincident with the ordinate axis of the plot.

The source was located at  $20^\circ$  of elevation and  $80^\circ$  of azimuth with a SNR of 0 dB. Figure 5 shows the normalized estimate, proportional to the likelihood. The estimated location of the source is  $80.83^\circ$  and  $20.18^\circ$ . It is important to remark that the normalized estimate (NMLM) performs like a high resolution procedure and its accurateness in the source location requires high density grid to scan for the maximum.

The Capon estimate for the power level of the background noise, with the spectral subtraction indicated previously can be viewed in Figure 6, where it is evident the similarity with the original one.

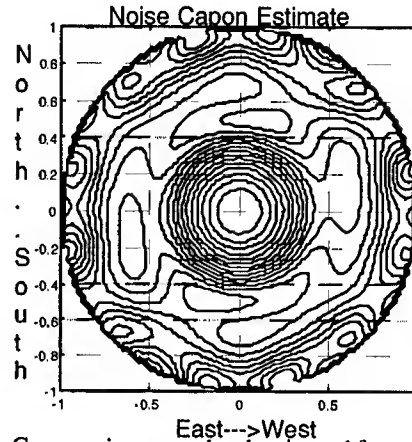


Figure 6. Capon noise power level estimated from the spectral subtraction procedure described in the text.

## VI. CONCLUSIONS

It has been shown what is the relationship between spectral density estimators and source detection in colored noise. At the same time, the interest of the normalized Capon estimate has been proven to be the natural 1-d or 2-d spectral density estimate. In fact, this density estimate, for any 2-d scenarios, is superior to the available procedures of Periodogram (low resolution) and Music (high complexity and order uncertainty).

## VII. REFERENCES

- [1] J. Capon. "High resolution frequency-wavenumber spectrum analysis". Proc. IEEE, Vol. 57, pp. 1408-1418, Aug. 1969.
- [2] P. Stoica, H. Li, J. Li. "Amplitude estimation of sinusoidal signals: Survey, new results, and an application". IEEE-SP Trans., Vol. 48, pp. 338-352, Feb. 2000.
- [3] M.A. Lagunas et al. "Maximum likelihood filters in spectral estimation". Signal Processing, EURASIP. Special issue in major trends in spectral estimation. Volume 10, No. 1, pp. 19-35. January 1986.
- [4] K.L. Bell, Y. Ephraim, L. Van Trees. "A bayesian approach to robust adaptive beamforming". IEEE SP Trans. Vol. 48, No. 2, pp. 386-398, Feb. 2000.

# PARAMETER ESTIMATION: THE AMBIGUITY PROBLEM

V. Lefkaditis, A. Manikas

Department of Electrical and Electronic Engineering,  
Imperial College of Science, Technology and Medicine.  
London, SW7 2BT, UK.

## ABSTRACT

In a statistical signal processing parameter estimation problem, ambiguities are generated when there is not a *one to one* mapping between the object space and the measurement space. In this paper, the ambiguity problem is investigated using differential geometry concepts and a theoretical framework is proposed for the classification, identification and calculation of ambiguities, associated with two application areas, that is in the array processing and in the harmonic retrieval problem.

## 1. INTRODUCTION

Much recent effort has been devoted to the study of ambiguities which is a well known research problem in parameter estimation applications. However, existing investigations have been restricted to the array processing area and are mainly concerned with four aspects of ambiguities. The first is the identification of array geometries free of ambiguities of up to a certain rank ([1], [2]). The second is concerned with the classification and estimation of manifold ambiguities for both linear and planar array geometries ([3], [4]), while the third studies the effects of sensor failure on the ambiguous behaviour of linear arrays [5]. The final one is related to resolving manifold ambiguities for non uniform linear array geometries [6], with all these four aspects providing also four independent lines of thought.

In this paper, the ambiguity problem is investigated from the prism of a general parameter estimation problem and a novel framework is proposed. This framework is appropriate for any parameter estimation problem provided that the parameter of interest  $p \in \Omega$  is mapped into a vector  $\underline{a}(p) \in \mathbb{C}^N$  (i.e.  $p \mapsto \underline{a}(p)$ ), with  $\underline{a}(p) \forall p \in \Omega$  being a curve having a hyperhelical shape. The modelling, which is based on differential geometry properties of hyperhelical curves, is presented in Section 2, along with some important concepts and definitions. In the same section, by partitioning a hyperhelical curve into uniform and non uniform segments, two classes of ambiguities are identified and an algorithm for estimating ambiguous sets of parameters is proposed in a general framework. Then in Sections 3 and 4 the application of this general framework to handle the array processing and the harmonic retrieval problems, respectively, is presented along with two representative examples. Finally, in Section 5 the paper is concluded.

## 2. THEORETICAL FRAMEWORK FOR THE ESTIMATION OF AMBIGUITIES

Let  $\underline{x}(t)$  be the observation signal in a parameter estimation problem, where the parameter of interest  $p \in \Omega$ , is mapped into a vector  $\underline{a}(p)$ , having the following general form:

$$\underline{a}(p) = \exp(-j\pi(\underline{v}h(p) + \underline{u})) \quad (1)$$

where  $\underline{v}, \underline{u}$  are constant  $N$ -dimensional real vectors and  $h(p)$  is a real function satisfying the following condition:

$$\left\| \frac{d\underline{a}(p)}{dp} \right\| = \frac{dh(p)}{dp} \quad \forall p \in \Omega$$

or

$$\left\| \frac{d\underline{a}(p)}{dp} \right\| = -\frac{dh(p)}{dp} \quad \forall p \in \Omega \quad (2)$$

Furthermore, if the observation signal  $\underline{x}(t)$  can be modelled as a function of  $\underline{a}(p)$ , as follows:

$$\underline{x}(t) = \sum_{i=1}^M \underline{a}(p_i) \cdot m_i(t) + \underline{n}(t) \quad (3)$$

or, in a matrix format,

$$\underline{x}(t) = \underline{A}(p) \cdot \underline{m}(t) + \underline{n}(t) \quad (4)$$

then the ambiguity problem arises when two or more vectors  $\underline{a}(p)$  are linearly dependent, leading to a rank deficient matrix  $\underline{A}(p)$ . In Equation (4),  $\underline{m}(t) = [m_1(t), m_2(t), \dots, m_M(t)]^T$  is a vector signal,  $\underline{n}(t)$  denotes the noise effects, and  $\underline{A}(p) = [\underline{a}(p_1), \underline{a}(p_2), \dots, \underline{a}(p_M)]$ .

It can be proven that the locus of the vector  $\underline{a}(p)$  given by Equation (1), over the parameter space  $\Omega$  ( $\forall p \in \Omega$ ), is a curve in  $N$ -dimensional complex space, having a hyperhelical shape. The advantages of having hyperhelical curves are numerous. The most important is that their shape and properties can be described by a set of constant curvatures which can be analytically estimated [7].

Since the locus of  $\underline{a}(p)$  is a curve embedded in an  $N$ -dimensional complex space, the *arc length*  $s$  is a much more natural way of parametrising a curve as compared to  $p$ , representing the actual length of a segment of the curve. If  $\underline{a}(p)$  is described by Equation (1) then the arc length  $s$  is related to the parameter of interest  $p$  via the following expression:

$$s(p) = \pi \|\underline{v}\| (h(p) - h(0)) \quad (5)$$

Note that the total length  $l_v$  of the curve is called the *manifold length*.

Although the number of ambiguous sets is infinite, it can be proven that if one ambiguous set of arc lengths is identified, then by simple rotation, an infinite number of ambiguous sets can be generated. For that reason the concept of the *ambiguous generator set* (representing all these rotated ambiguous sets) and its corresponding rank, called *rank of ambiguity* were introduced in [3] as follows:

**Ambiguous generator set:**

An ordered set  $\underline{s} = [0, s_1, \dots, s_{M-1}]^T$  of  $M$  arc lengths, where  $2 \leq M \leq N$ , is said to be an *ambiguous generator set of arc lengths* if and only if:

- a) All the elements of the set but the first element are non-zero,
- b) The rank of the  $N \times M$  matrix  $\mathbb{A}(\underline{s})$  with columns the manifold vectors associated with the elements of the set is less than  $M$ , i.e.  $\text{rank}(\mathbb{A}(\underline{s})) = \rho < M$  and
- c) For any subset  $\underline{s}_i$  of  $k$  elements of  $\underline{s}$  with  $\rho \leq k < M$ , the rank of  $\mathbb{A}(\underline{s}_i)$  is equal to  $\rho$ .

**Rank of Ambiguity:** The rank of the matrix  $\mathbb{A}(\underline{s})$  is called *rank of ambiguity*  $\rho$  of the set  $\underline{s}$ .

The ambiguous generator sets are divided in two different classes according to the way the hyperhelical curve is partitioned. That is the uniform class and the non uniform class. However, the non uniform class exists if and only if the vector  $\underline{v}$  is a *symmetric* vector, i.e. if the elements of  $\underline{v}$  are symmetric with respect to their centroid, i.e.  $\sum (\underline{v}^i) = 0, \forall i = \text{odd}$ , where  $\underline{v} = \underline{v} + b\underline{1}$ , with  $b$  a real number.

In the case of the uniform class of ambiguities, a set of arc lengths  $\underline{s}_{i,j}$  (that partitions the hyperhelix uniformly) can be formed by the following equation:

$$\underline{s}_{i,j} = q \left[ 0, \frac{1l_v}{|v_i - v_j|}, \dots, \frac{cl_v}{|v_i - v_j|} \right]^T \quad (6)$$

where  $v_i, v_j$  are elements of the vector  $\underline{v} \forall i, j$ ,  $c$  is an integer number and

$$q = \frac{2}{h(p_{\max}) - h(0)} \quad (7)$$

The existing sets of the above form provide the corresponding ambiguous generator sets,  $\forall i, j$ .

In the case of the non uniform class of ambiguities, the hyperhelical curve is partitioned to a number of non uniform segments according to the roots of

$$\text{Tr}(\mathbb{C} \exp\{s \mathbb{C}\}) = 0 \quad \forall s \in [0, l_v) \quad (8)$$

where  $\mathbb{C}$  is a  $d \times d$  matrix, known as the *Cartan matrix* and defined as:

$$\mathbb{C} \stackrel{\text{def}}{=} \begin{bmatrix} 0 & -\kappa_1 & 0 & 0 & \dots & 0 \\ \kappa_1 & 0 & -\kappa_2 & 0 & \dots & 0 \\ 0 & \kappa_2 & 0 & -\kappa_3 & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \kappa_{d-2} & 0 & -\kappa_{d-1} \\ 0 & 0 & \dots & 0 & \kappa_{d-1} & 0 \end{bmatrix} \quad (9)$$

with  $\kappa_i$  being the  $i^{\text{th}}$  curvature [7], and

$$d = \begin{cases} 2N - k & \text{if no element of } \underline{v} \text{ is at the centroid of } \underline{v} \\ 2N - k - 1 & \text{otherwise} \end{cases}$$

with  $k$  representing the number of elements of  $\underline{v}$  which have symmetrical pairs with respect to its centroid.

Based on Equations (6) and (8), an algorithm for estimating the *ambiguous generator sets* and the associated *rank of ambiguity* is presented below in a compact step format:

**Estimation Algorithm:**

**STEP 1:** Calculate the length  $l_v$  of the hyperhelical curve and the vector  $\hat{\underline{v}}$  which is the Kronecker difference  $\underline{v} \ominus \underline{v}$  with all the elements that are smaller than one eliminated. Also, if the vector  $\underline{v}$  is symmetric, calculate the Cartan matrix  $\mathbb{C}$ .

**STEP 2:** For each of the elements of the vector  $\hat{\underline{v}}$ , create the corresponding set  $\underline{s}_{i,j}$ , by partitioning the hyperhelix uniformly, using Equation (6). These sets will provide the ambiguous generator sets belonging to the uniform class. If the vector  $\underline{v}$  is also symmetric then calculate the roots of Equation (8) and form an extra set  $\underline{s}_{i,j}$ . This set will provide the ambiguous generator sets belonging to the non uniform class.

**STEP 3:** For each of the sets formed in Steps 2, identify the ambiguous generator sets based on the following rules:

rule-a) If  $\underline{s}_{i,j}$  is unique, then  $\binom{k}{N-1}$  ambiguous generator sets can be produced which are all the possible subsets of  $N$  elements of the set with their first element zero and  $N-1$  non-zero elements. Their rank of ambiguity is equal to  $N-1$ .

rule-b) If  $\underline{s}_{i,j}$  is not unique, then all subsets of  $\underline{s}_{i,j}$  with their first element 0 and with length 2, 3, ... up to  $\min(N, k+1)$  must be considered. These subsets are classified as ambiguous generator sets if the three conditions of the ambiguous generator set definition are satisfied [3]. Furthermore, for each ambiguous generator set  $\underline{s}_{i,j}$ , the rank of ambiguity  $\rho_{ij}$  is determined.

### 3. ARRAY PROCESSING

In an azimuth direction finding system employing planar arrays, the signal  $\underline{x}(t)$  is modelled as in Equation (3), with  $\underline{m}(t)$  being a baseband signal-vector and  $\underline{n}(t)$  the noise vector. The parameters of interest ( $\theta$  azimuth,  $\phi$  elevation) are mapped to the array manifold vector as follows:

$$\underline{a}(\theta, \phi) = \exp(-j\pi(\underline{r}_x \cos\theta + \underline{r}_y \sin\theta) \cos\phi) \quad (10)$$



where  $[r_x, r_y, 0]$  denote the locations of the sensors in half-wavelengths and  $\theta \in [0, 2\pi)$ ,  $\phi \in [0, \frac{\pi}{2})$ .

If Equation (10) is compared with Equation (1), it can be seen that the  $\phi$ -curves ( $\theta = \text{constant}$ ) are hyperhelical curves with  $p = \phi$ ,  $\underline{v} = r_x \cos \theta + r_y \sin \theta$  and  $\underline{u} = 0$ . However,  $\theta$ -curves ( $\phi = \text{constant}$ ) are not hyperhelices. Therefore ambiguities associated only with the  $\phi$ -curves, (i.e. elevation angles), can be estimated by using the proposed approach. Furthermore, based on *cone-angle parametrisation* ( $\alpha, \beta$ ) [8] which is an alternative parametrisation to  $(\theta, \phi)$ , the array manifold vector can be rewritten as follows:

$$\underline{a} = \exp\left(-j\pi\left(\underline{R}(\Theta)\cos\alpha + \underline{R}(\Theta + \frac{\pi}{2})\cos\beta\right)\right) \quad (11)$$

$\forall \alpha, \beta \in [0^\circ, 180^\circ)$ , where  $\begin{cases} \underline{R}(\Theta) = r_x \cos \Theta + r_y \sin \Theta \\ \Theta \text{ is the rotation of the } x\text{-}y \text{ frame} \end{cases}$

matching Equation (1) and thus both  $\alpha$ -curves ( $\beta = \text{constant}$ ) and  $\beta$ -curves ( $\alpha = \text{constant}$ ), are hyperhelical curves. For instance, for the  $\alpha$ -curves,  $p = \alpha$ ,  $\underline{v} = \underline{R}(\Theta)$  and  $\underline{u} = \underline{R}(\Theta + \frac{\pi}{2})\cos\beta$ .

From the previous discussion and modelling it is obvious that ambiguities associated with the  $\phi$ ,  $\alpha$  and  $\beta$ -curves can be estimated using the proposed approach, while the case of constant elevation and different azimuths ( $\theta$ -curves), remains an open problem.

**Example 1:** Consider a planar array with sensor locations given by the following matrix, in half-wavelengths:

$$[r_x, r_y, 0] = \begin{bmatrix} -1.7, & -1.5, & 0, & 0, & 1.5, & 1.7 \\ 2.2, & -2.5, & -2, & 2, & 2.5, & -2.2 \\ 0, & 0, & 0, & 0, & 0, & 0 \end{bmatrix}^T$$

For the  $\phi$ -curve of the array manifold corresponding to  $\theta = 85^\circ$  (say), the manifold length  $l_v(85^\circ)$  is equal to 34.4330. If the proposed method described in the previous section is applied to calculate the ambiguous generator sets associated with this  $\phi$ -curve, then the following ambiguous generator sets in arc lengths are estimated:

$$\underline{\Sigma} = \begin{bmatrix} 0, & 6.5681, & 13.1362, & 19.7043, & 26.2725, & 32.8406 \\ 0, & 7.3816, & 14.7633, & 22.1449, & 29.5265, & 0 \\ 0, & 7.4633, & 14.9267, & 22.3900, & 29.8534, & 0 \\ 0, & 8.5318, & 17.0635, & 25.5953, & 34.1271, & 0 \end{bmatrix}$$

with corresponding rank of ambiguity  $\underline{\rho} = [5, 4, 4, 4]^T$ .

Furthermore, if the ambiguous generator sets of the above array are estimated for every  $\theta$ , then *ambiguous generator lines* will be formed. Figure 1 shows the set of ambiguous generator lines associated with the first row of the matrix  $\underline{\Sigma}$ . In the same figure the locus of the manifold lengths of every  $\phi$ -curve ( $l_v(\theta), \forall \theta$ ) is also shown as a dashed line. The intersection of a line from the origin with a set of ambiguous generator lines provides an ambiguous generator set of directions and in Figure 1 the

symbols ( $\bullet$ ) show an ambiguous generator set for  $\theta = 85^\circ$  (first row of  $\underline{\Sigma}$ ).

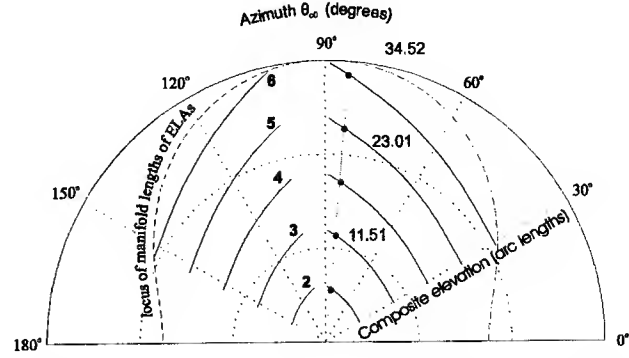


Figure 1: The set of ambiguous generator lines of rank-5 associated with the first row of  $\underline{\Sigma}$ .

## 4. HARMONIC RETRIEVAL PROBLEM

Consider a signal  $x(t)$  which is a sum of  $M$  complex sinusoids, with unknown amplitudes  $\underline{m} = [m_1, \dots, m_i, \dots, m_M]^T$  and unknown frequencies  $\underline{f} = [f_1, \dots, f_i, \dots, f_M]^T$  which are to be estimated. The frequencies are normalised with respect to a known maximum frequency  $f_s$  which implies that:

$$0 \leq f_i < 1 \quad \forall i \quad (12)$$

The signal is assumed to be contaminated with additive white Gaussian noise,  $n(t)$ .

Let us also consider that over an observation interval  $T_{\text{obs}}$ , the signal  $x(t)$  is sampled at a non-uniform rate, with the number of samples,  $N$ , satisfying the following condition:

$$M \leq N \leq \left\lfloor \frac{T_{\text{obs}}}{T_s} \right\rfloor \quad (13)$$

where  $T_s$  is defined as  $1/f_s$ . This implies that the signal  $x(t)$  is sampled at times  $t_1, t_2, \dots, t_N \in \mathbb{R}$ , which are normalised with respect to  $T_s$ . By defining the  $N \times 1$  vector  $\underline{t} = [t_1, \dots, t_N]^T \in \mathbb{R}^N$ , which for uniform sampling at the Nyquist rate becomes  $[1, 2, \dots, N]^T$ , the data sequence over the observation interval  $T_{\text{obs}}$ , can be modelled as follows:

$$\underline{x} = \sum_{i=1}^M m_i \exp(j2\pi \underline{t} f_i) + \underline{n} \quad (14)$$

or in matrix format  $\underline{x} = \mathbb{A}(\underline{f}) \underline{m} + \underline{n}$  (15) where  $\mathbb{A}(\underline{f}) = \exp(j2\pi \underline{t} \underline{f}^T)$  is an  $N \times M$  matrix and  $\exp(\cdot)$  denotes element by element exponential.

Let us now define as the *frequency manifold vector* the following  $N \times 1$  vector:

$$\underline{a}(\underline{f}) = e^{j2\pi \underline{t} \underline{f}} \quad \forall \underline{f} \in [0, 1) \quad (16)$$

This vector is of the form of Equation (1) with  $p = \underline{f}$ ,  $\underline{v} = \underline{t}$  and  $\underline{u} = 0$ , and its locus is a hyperhelical curve. Thus the proposed algorithm of Section 2, can be employed to estimate the sets of

ambiguous generator frequencies, as can be seen by the following example.

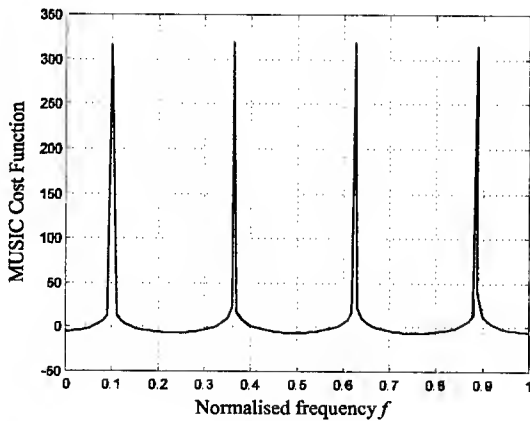
**Example 2:** Consider a signal which is the sum of three sinusoids of normalised unknown frequencies 0.3632, 0.6263 and 0.8895, plus additive white Gaussian noise. If the signal is sampled over a normalised observation interval at times  $\underline{t} = [0, 1.3, 2.5, 3.8]^T$ , then three ambiguous generator sets of frequencies (and their associated *rank of ambiguity*) can be found by the proposed algorithm. These are:

$$\mathbb{F} = \begin{bmatrix} 0, & 0.2632, & 0.5263, & 0.7895 \\ 0, & 0.4000, & 0.8000, & 0 \\ 0, & 0.2868 & 0.5008 & 0.7934 \end{bmatrix} \quad \rho = \begin{bmatrix} 3 \\ 2 \\ 3 \end{bmatrix} \quad (17)$$

with the manifold length of the associated hyperhelix being equal to:

$$l_v = 2\pi\|\underline{t}\| = 29.7242 \quad (18)$$

Note that because the vector  $\underline{t}$  is a symmetric vector, (i.e.  $b = -1.9$  and  $\sum(\tilde{t}^i) = 0, \forall i = \text{odd}$ ), both classes of ambiguities exist and in particular the first two rows (sets) of  $\mathbb{F}$  belong to the 'uniform' class while the last row of  $\mathbb{F}$  to the 'non-uniform' class of ambiguities. If the MUSIC algorithm is used to estimate the unknown frequencies, four frequencies will be provided rather than three and this is illustrated in Figure 2. It is clear that the estimated frequencies are  $\underline{f} = [0.1, 0.3632, 0.6263, 0.8895]$ . It is obvious that by subtracting 0.1 from the elements of  $\underline{f}$  the first row of the matrix  $\mathbb{F}$  is obtained, indicating that the estimated frequencies are 'ambiguous'.



**Figure 2:** MUSIC spectrum for the set of samples of the Example 2. The true frequencies are [0.3632, 0.6263, 0.8895] and the frequencies estimated by MUSIC are [0.1, 0.3632, 0.6263, 0.8895].

From the above example it can be seen that if the unknown frequencies do not correspond to any of the ambiguous generator sets (rows of matrix  $\mathbb{F}$ ), then they can be estimated unambiguously, even if the sampling rate is lower than Nyquist. It is also obvious that by using any set of  $N$  non-uniform samples

which provides a number of ambiguous generator sets with minimum rank of ambiguity  $\rho$ , we can unambiguously resolve any  $\rho - 1$  frequencies. Note that the maximum number of unambiguously estimated frequencies for a set of  $N$  samples is achieved when the minimum rank of ambiguity  $\rho$  is equal to  $N - 1$ .

## 5. CONCLUSIONS

In this paper, the ambiguity problem has been investigated and a generalised framework has been proposed for calculating the ambiguous sets of parameters, based on the hyperhelical parametrisation of the manifold vectors. The proposed framework was supported by two representative examples, one associated with the harmonic retrieval problem and the other with the array processing area.

## REFERENCES

- [1] K. C. Tan and Z. Goh, "A Detailed Derivation of Arrays Free of Higher Rank Ambiguities", *IEEE Trans. on Signal Processing*, Vol. 44, No. 2, pp. 351-359, February 1996.
- [2] K. C. Tan, S. S. Goh and E. C. Tan, "A Study of the Rank-Ambiguity Issues in Direction-of-Arrival Estimation", *IEEE Trans. on Signal Processing*, Vol. 44, No. 4, pp. 880-887, April 1996.
- [3] A. Manikas, C. Proukakis, "Modeling and Estimation of Ambiguities in Linear Arrays", *IEEE Trans. on Signal Processing*, Vol. 46, No. 8, pp. 2166-2179, August 1998.
- [4] A. Manikas, C. Proukakis and V. Lefkaditis, "Investigative Study of Planar Array Ambiguities Based on 'Hyperhelical' Parametrisation", *IEEE Trans. on Signal Processing*, Vol. 47, No. 6, pp. 1532-1541, June 1999.
- [5] V. Lefkaditis and A. Manikas, "Investigation of Sensor Failure with respect to Ambiguities in Linear Arrays", *IEE Electronics Letters*, Vol. 35, No. 1, pp. 22-23, January 1999.
- [6] Y. I. Abramovich, N. K. Spencer and A. Y. Gorokhov, "Resolving Manifold Ambiguities in Direction-of-Arrival Estimation for Non uniform Linear Antenna Arrays", *IEEE Trans. on Signal Processing*, Vol. 47, No. 10, pp. 2629-2643, October 1999.
- [7] I. Dacos and A. Manikas, "Estimating the Manifold Parameters of One-dimensional Arrays of Sensors", *Journal of the Franklin Institute*, Vol. 332B, No. 3, pp. 307-332, 1995.
- [8] H. R. Karimi and A. Manikas, "Cone-Angle Parametrization of the Array Manifold in DF Systems", *Journal of the Franklin Institute*, Vol. 335B, No. 2, pp. 375-394, 1998.

*Acknowledgment:* V. Lefkaditis gratefully thanks the Greek State Scholarships Foundation (IKY) for the scholarship they have awarded him.

# ON MULTIWINDOW ESTIMATORS FOR CORRELATION

Alfred Hanssen

University of Tromsø, Department of Physics  
Electrical Engineering Group, N-9037 Tromsø, Norway  
E-mail: alfred@phys.uit.no.

## ABSTRACT

We have examined the bias and variance properties of a recently suggested class of multiwindow estimators for autocorrelation functions (ACF). The derived exact expression for the bias is valid for any amplitude distribution, while the derived exact result for the variance is valid for zero-mean Gaussian processes. We show that the multiwindow ACF estimator has undesirable bias properties and inferior variance properties compared to the standard ACF estimator. The reason is that the correlation properties of the windows contribute directly to the ACF estimator and its statistical moments. The lesson to be learned is that what is good for spectral estimators is not necessarily good for correlation estimators.

## 1. INTRODUCTION

Applications often require that accurate estimates of the Autocorrelation Function (ACF) of some underlying stochastic process is known. It is in general a difficult task to estimate the ACF from data, because often only short sampled data segments are available. This introduces a certain estimation bias and variance in the estimates. A natural goal is to reduce the bias and/or variance as much as possible.

In this paper, we examine the bias and variance properties of a recently suggested class of ACF estimators [2, 3] that is based on Thomson's multiwindow spectral estimators [7]. The statistical properties of the novel estimators will be compared to those of a classical ACF estimator.

The basic definition of an ACF for a wide sense stationary stochastic process  $X(t)$  is (e.g., Ref. [4])

$$R_{XX}(\tau) = E[X(t)X(t + \tau)], \quad (1)$$

where  $E[\cdot]$  is the expectation operator, and  $\tau$  is the time lag.

The Power Spectral Density (PSD)  $S_{XX}(f)$  of a wide-sense stationary process  $X(t)$  is related to the

ACF through the Wiener-Khinchine theorem as

$$S_{XX}(f) = \mathcal{F}\{R_{XX}(\tau)\}, \quad (2)$$

where  $\mathcal{F}\{\cdot\}$  is the Fourier transform.

## 2. MULTIWINDOW PSD ESTIMATORS

The fairly recent multiwindow (MW) non-parametric estimator for power spectral densities (PSD) [7, 4, 5] can be seen as a variation or an extension of the windowed periodogram technique. In this method, one applies a sequence of orthogonal data windows that obey some optimality criterion, to form a sequence of direct windowed PSD estimates. The windowing reduces the spectral leakage, as is well-known from classical spectral estimation. By forming a weighted average of the individual spectral estimates, we are simultaneously able to reduce the estimation variance.

### 2.1. Discrete Prolate Spheroidal Sequences

Thomson [1982] proposed to apply some stringent optimality criteria when selecting data tapers. He suggested to consider tapers that maximizes the "spectral concentration", or the energy contained in the mainlobe relative to the total energy of the taper. One therefore seeks the taper  $v[n]$  with a discrete Fourier transform  $V(f)$ , that maximizes the window energy ratio

$$\lambda = \frac{\int_{-f_B}^{f_B} |V(f)|^2 df}{\int_{-1/2}^{1/2} |V(f)|^2 df} \quad (3)$$

where  $f_B$  is the wanted resolution half-bandwidth (a design parameter) of the taper. An ideal taper would therefore have  $\lambda \simeq 1$  and  $f_B$  as small as possible (but note that  $f_B > 1/N$ ). (Note also that we use  $\Delta t = 1$  in this chapter to simplify the notation.)

Expressing  $V(f)$  by its discrete Fourier transform,  $V(f) = \sum_{n=0}^{N-1} v[n] \exp(-j2\pi fn)$  and maximizing the above functional with respect to  $v[n]$ ,

Slepian [1978] showed that the optimal taper  $\mathbf{v} = [v[0], v[1], \dots, v[N-1]]^T$  obeys the eigenvalue equation

$$\mathbf{A}\mathbf{v} = \lambda\mathbf{v} \quad (4)$$

where the matrix  $\mathbf{A}$  has elements  $[\mathbf{A}]_{nm} = \sin[2\pi f_B(n-m)]/[\pi(n-m)]$ , for  $n, m = 0, 1, \dots, N-1$ . Note that (4) is an  $N$ -dimensional eigenvector/eigenvalue problem, thus giving  $N$  eigenvector/eigenvalue pairs,  $(\mathbf{v}_k, \lambda_k)$ , where  $k = 0, 1, \dots, N-1$ . The interpretation is thus that we obtain a sequence of orthogonal tapers (eigenvectors),  $\mathbf{v}_k$ , each with a corresponding spectral concentration measure  $\lambda_k$ . The first taper  $\mathbf{v}_0$  has a spectral concentration  $\lambda_0$ . Then,  $\mathbf{v}_1$  maximizes the ratio in (3) subject to being orthogonal to  $\mathbf{v}_0$ , and with  $\lambda_1 < \lambda_0$ . Continuing, we can thus form up to  $N$  orthogonal tapers  $\mathbf{v}_0, \mathbf{v}_1, \dots, \mathbf{v}_{N-1}$ , with  $0 < \lambda_{N-1} < \lambda_{N-2} < \dots < \lambda_0 < 1$ . Only tapers with  $\lambda_k \simeq 1$  can be applied, since  $\lambda_k \ll 1$  implies a large undesirable leakage. It is usually safe to apply  $K = 2Nf_B$  tapers [Percival and Walden, 1993, pp. 334-335].

There are many ways to form weighted averages over the windowed data. We may therefore write a general MW PSD estimate as

$$\hat{S}_{MT}(f) = \sum_{k=0}^{K-1} \alpha_k \hat{S}_{MT}^{(k)}(f) \quad (5)$$

where the "eigenspectrum" of order  $k$  is defined by

$$\hat{S}_{MT}^{(k)}(f) = \left| \sum_{n=0}^{N-1} v_k[n] x[n] \exp(-j2\pi f n) \right|^2 ; \quad |f| \leq 1/2 \quad (6)$$

where  $v_k[n]$  denotes the elements of DPSS-taper of order  $k$ , and  $\alpha_k$  is a weight factor for eigenspectrum no.  $k$ .

The three "standard" weight coefficients are (i) Uniform weighting,  $\alpha_k = 1/K$ ,  $k = 0, \dots, K-1$ , (ii) Eigenvalue weighting,  $\alpha_k = \lambda_k / \sum_{k=0}^{K-1} \lambda_k$ , and (iii) Inverse eigenvalue weighting,  $\alpha_k = 1/\lambda_k \left( \sum_{k=0}^{K-1} \lambda_k \right)$ .

### 3. MULTIWINDOW ACF ESTIMATORS

The rationale behind the MW-ACF estimator is the following. If a spectral estimate has been derived from a high-quality estimator like the MW spectral estimator, a direct application of Wiener-Khinchine's theorem should produce a high-quality ACF estimator. Thus, the MW-ACF estimator is given by

$$\hat{R}_{MW}[m] = \mathcal{F}^{-1} \left\{ \hat{S}_{MW}(f) \right\} \quad (7)$$

where  $\mathcal{F}$  denotes the Fourier transform, a hat denotes an estimator, and subscript  $MW$  denotes multiwindow.

Assume that the data available from a single realization  $x(t)$  of a process  $X(t)$  are

$$x[n] \equiv x(n\Delta t) \quad ; \quad n = 0, 1, \dots, N-1 \quad (8)$$

where  $\Delta t$  is the sampling interval.

It is easy to show that Eq. (7) leads to an estimator of the form [2, 3]

$$\hat{R}_{MW}[m] = \frac{1}{K} \sum_{k=0}^{K-1} \hat{R}_{MW}^{(k)}[m] \quad (9)$$

where

$$\hat{R}_{MW}^{(k)}[m] = \sum_{n=0}^{N-1-|m|} v_k[n] v_k[n+|m|] x[n] x[n+|m|] \quad (10)$$

Here,  $x[n]$  is the datum at time step  $n$ ,  $v_k[n]$ ;  $n = 0, 1, \dots, N-1$  are the components of taper no.  $k$ , and  $K$  is the number of tapers applied in the formation of the ACF-estimate. Usually, one chooses  $K \ll N$  to avoid excessive leakage from the tapers.

The tapers  $v_k[n]$  that maximizes the energy contained in the main lobe, subject to a designer specified half-bandwidth, are the so-called Slepian sequences, or Discrete Prolate Spheroidal Sequences (DPSS) [6, 7]. These tapers cannot be written in a closed form, but are rather defined as a solution of an eigenvalue, eigenvector problem [6, 7, 4]. Recently, a simpler set of orthonormal tapers were introduced by [5]. These tapers are commonly referred to as "sinusoidal tapers" due to their mathematical definition. The sinusoidal tapers are approximations to tapers that minimize the local bias, subject to being orthonormal in sample space.

Note that the classical biased ACF-estimator (the "standard" ACF-estimator [4]) is derived from (9) simply by choosing  $K = 1$  and  $v_0[n] = 1/\sqrt{N}$ ;  $n = 0, 1, \dots, N-1$ .

#### 3.1. Expectation Value

It is straightforward to evaluate the expectation value of the multiwindow ACF estimator. We found that

$$\frac{E \left\{ \hat{R}_{MW}[m] \right\}}{R_{XX}[m]} = Q[m], \quad (11)$$

where

$$Q[m] = \sum_{k=0}^{K-1} \alpha_k \rho_k[m]$$

where  $\rho_k[m] = \sum_{n=0}^{N-1-|m|} v_k[n] v_k[n+|m|]$  is the deterministic correlation function for data window no.  $k$ .

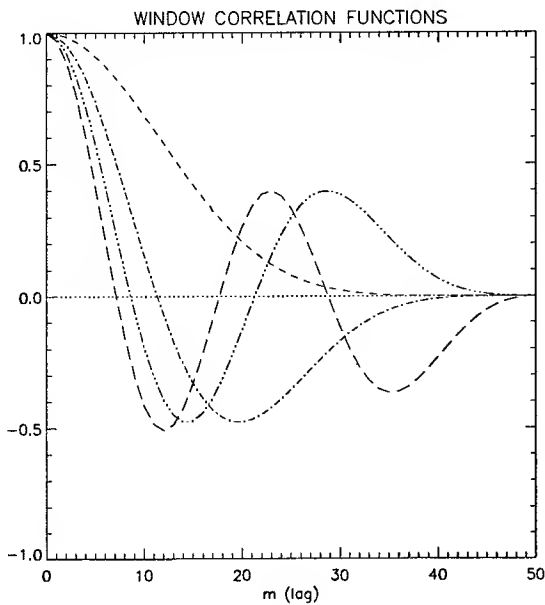


Figure 1: The window correlation autocorrelation  $\rho_k[m]$  for window orders  $k = 0$ : short dashes ;  $k = 1$ : dash-dot ;  $k = 2$ : dash-dot-dot-dot ; and  $k = 3$ : long dashes.

Thus, the expectation value of the MW-ACF estimator is governed by a weighted average of the correlation functions of the  $K$  individual taper sequences. Note that this result is exact, and that no assumptions were made about the amplitude distribution.

The result in (11) is very important, because it shows that any of the three standard weightings will cause severe problems for the estimator. For  $K > 1$ , we will end up with lag-regions where  $Q[m] < 0$ , and the estimate will in general be severely biased. It is easy to understand why this is so by examining the window correlation functions order by order. In Fig. 1 we show the four lowest order window correlation functions  $\rho_k[m]$  for  $k = 0, 1, 2, 3$  for DPSS windows with  $N = 50$  and  $Nf_B = 3$ . We see that all correlation functions are decaying, and that the higher the order, the more oscillations is evident.

In Fig. 2 we show the quantity  $Q[m]$  (see Eq. (11)), which is the expectation value normalized by the true ACF for each lag. We have shown  $Q[m]$  for  $K = 1, 2, 3$ , and 4 tapers using a uniform weighting of the individual windowed ACF estimates. The full line is the result for the classical (biased) ACF-estimator for comparison, the short-dashed curve is the MW-ACF case for  $K = 1$ , the dash-dot curve is  $K = 2$ , the dash-dot-dot-dot curve is  $K = 3$ , and the long-dashed curve is  $K = 4$ . Note that  $Q[m]$  with uniform weighting in general

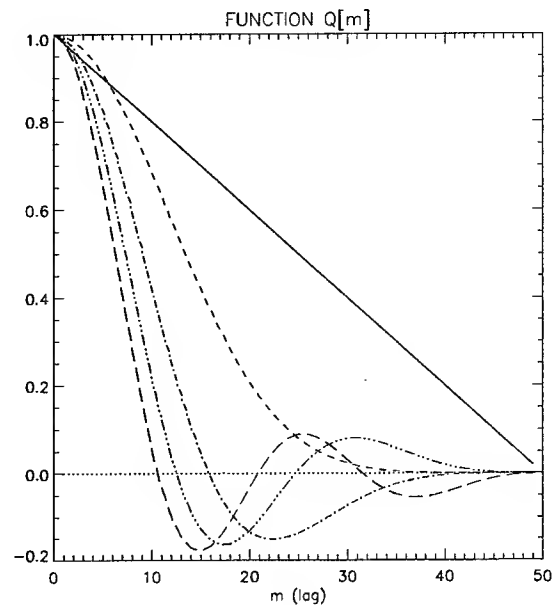


Figure 2: The function  $Q[m]$ . Classical ACF-estimator: full line; MW-ACF with  $K = 1$ : short dashes ;  $K = 2$ : dash-dot ;  $K = 3$ : dash-dot-dot-dot ; and  $K = 4$ : long dashes.

exhibits  $K - 1$  zeros, where  $K$  is the number of tapers.

It is important to notice that the MW-ACF estimators in general introduce a significant bias. When  $K > 1$ , we see that there exist lag-ranges where the expected value of the estimator has an incorrect sign. In general, the range of lags where one have some degree of confidence in the expectation value of the MW-ACF estimator, diminishes as  $K$  increases.

### 3.2. Variance

In general, it is impossible to evaluate the variance of the MW-ACF estimator. This is because the result will depend explicitly on the probability density function of the process amplitude. By making certain standard assumptions, however, we are able to derive some expressions that shed some light on the variability of the MW-ACF estimator.

By assuming the process to be a zero-mean real-valued Gaussian stochastic process, it is possible to show that

$$\text{var} \{ \hat{R}_{MW}[m] \} = \sum_{k=0}^{K-1} \sum_{l=0}^{K-1} \alpha_k \alpha_l \{ F(k, l; m) + G(k, l; m) \} \quad (12)$$

where

$$F(k, l; m) = \sum_{n=0}^{N-1-|m|} \sum_{n'=0}^{N-1-|m|} v_k[n]v_k[n+|m|]v_l[n']v_l[n'+|m|]R_{XX}^2[n'-n]$$

and

$$G(k, l; m) = \sum_{n=0}^{N-1-|m|} \sum_{n'=0}^{N-1-|m|} v_k[n]v_k[n+|m|] \times v_l[n']v_l[n'+|m|]R_{XX}[n'-n-|m|]R_{XX}[n'-n+|m|]$$

The variance of the MW-ACF estimator is thus governed by the fourth-order properties of the taper sequences, but it also depends explicitly on the fourth-order properties of the true ACF of the process. Eq. (12) may of course be evaluated numerically for a given true ACF, but as it stands, this expression may seem of little use.

To simplify further, we now assume that the process is white and has a variance  $\sigma_X^2$ . Under these assumptions, we find that the variance has the form

$$\frac{\text{var} \{ \hat{R}_{MW}[m] \}}{\sigma_X^4} = \sum_{k=0}^{K-1} \sum_{l=0}^{K-1} \alpha_k \alpha_l \rho_{k,l}[m] \quad ; \quad m \neq 0 \quad (13)$$

where

$$\rho_{k,l}[m] = \sum_{n=0}^{N-1-|m|} v_k[n]v_l[n]v_k[n+|m|]v_l[n+|m|]$$

is a fourth order window correlation function involving the windows at two different orders  $k$  and  $l$ .

### 3.3. Related Quadratic Error Measures

A related important quadratic error measures that combines the bias and variance is the mean-squared error (MSE)

$$\text{mse} \{ \hat{R}_{MW}[m] \} = \text{var} \{ \hat{R}_{MW}[m] \} + B^2 \{ \hat{R}_{MW}[m] \} \quad (14)$$

where the bias is  $B \{ \hat{R}_{MW}[m] \} = (Q[m] - 1) R_{XX}[m]$ . The cumulative MSE up to lag  $m$

$$\text{cmse} \{ \hat{R}_{MW}[m] \} = \sum_{l=0}^m \text{mse} \{ \hat{R}_{MW}[l] \}. \quad (15)$$

may also be used to quantify the performance of the estimator.

## 4. NUMERICAL EXAMPLES

In the following numerical examples, we have applied the Discrete Prolate Spheroidal Sequences (DPSS) [6] as data windows. These windows maximize the window energy in the main lobe whose bandwidth is a user specified parameter. The DPSS's are not expressible on closed form, rather, they are the solution of the eigenproblem

$$\mathbf{A} \mathbf{v} = \lambda \mathbf{v} \quad (16)$$

where the elements of the matrix  $\mathbf{A}$  are given by  $A_{mn} = \sin[2\pi f_B(m-n)]/[\pi(m-n)]$ ,  $m, n = 0, 1, \dots, N-1$ , and  $f_B$  is the desired resolution half-bandwidth of the tapers. All examples shown are for data sets of length  $N = 50$ , and a bandwidth parameter of  $f_B = 3/N$ .

### 4.1. Autoregressive process of order one

Autoregressive processes of order one (AR(1)) are Gaussian, and have an ACF given by

$$R_{XX}[m] = \frac{\sigma^2}{1-a_1^2} (-a_1)^{|m|} \quad (17)$$

where  $a_1$  is the AR-parameter, and  $\sigma^2$  is the variance of the driving zero-mean Gaussian noise. In the example to follow, we have chosen the parameters  $a_1 = -0.5$  and  $\sigma^2 = 1$ .

In Fig. 2 we show the exact bias, variance, mean-squared error, and cumulative MSE for four different MW-ACF estimators, compared to the exact results for the classical ACF-estimator. The different line-styles has the same meaning as in Fig. 1.

We see that for  $K = 1, 2, 3$ , the peak value of the bias is lower for the MW-ACF than it is for the classical ACF estimator, whereas  $K = 4$  has a maximum bias that is larger than that of the classical estimator. It is very important to notice that the classical ACF estimator has a variance, MSE, and cumulative MSE that is *lower* than the those of the MW-ACF estimators, for the small time lags. Beyond some crossover lag, however, the variance, MSE, and cumulative MSE of the MW-ACF decreases drastically, and stays far below the corresponding values for the classical ACF estimator.

In [3] they estimated the cumulative MSE for AR(2) and MA(2)-data by means of a Monte Carlo simulation. Their simulation results are consistent with our exact results for the MW-ACF estimator. They did however not compare their results to that of the classical ACF estimator. We have found that also for their examples will the classical ACF estimator outperform the MW-ACF estimator for small time lags.

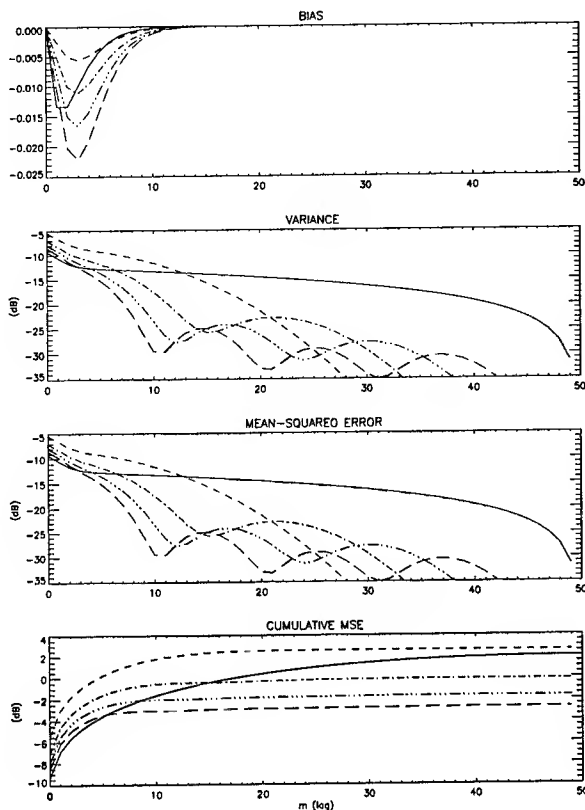


Figure 3: Exact bias, variance, MSE, and cumulative MSE for MW-ACF estimators and the classical ACF estimator for an AR(1)-process. Line-styles as in Fig. 1.

## 5. CONCLUSION

We derived an exact expression for the expectation value of the multiwindow ACF (MW-ACF) estimator valid for any amplitude distribution. Furthermore, we found two useful approximations for the estimator variance in the case of zero-mean Gaussian data, one for colored processes, and one for white noise. By comparing the bias, variance, mean-squared error (MSE), and cumulative MSE of the MW-ACF estimators with those of the classical estimator, we see that there is a trade-off between the variance reduction and the bias introduced by the tapering. For small time lags, the MW-ACF always exhibits a *larger* variance, MSE and cumulative MSE than that of the classical estimator. For the larger lags, the MW-ACF has a far *less* variance, MSE and cumulative MSE, but the expected value of the estimator can in turn be unacceptably large. In general, the MW-ACF estimator even induces an incorrect sign of the expected value for certain time-

lag intervals.

It is evident that only for very limited lag-ranges will the MW-ACF estimator be able to outperform the classical ACF estimator. We must therefore conclude that the statistical properties of the MW-ACF estimator are such that this estimator will be of limited importance for solving real world ACF estimation problems.

The reason for this peculiar behavior is that the correlation properties of the multiple windows becomes important for all moments of the MW-ACF estimator.

The lesson to be learned from this is that what is good for spectral estimation is not necessarily so for a correlation estimation - despite the existence of the Wiener-Khinchine theorem.

## REFERENCES

- [1] A. Hanssen, Multidimensional multitaper spectral estimation, *Signal Processing*, Vol. 58, pp. 327-332, 1997.
- [2] A. Hanssen and M. Finsrud, Multitaper estimators for the autocorrelation function, *Proc. 1998 Eighth IEEE DSP Workshop*, Utah, 1998.
- [3] L. T. McWhorter and L. L. Scharf, Multiwindow estimators of correlation, *IEEE Trans. Signal Processing*, Vol. 46, pp. 440-448, 1998.
- [4] D. B. Percival and A. T. Walden, *Spectral Analysis for Physical Applications: Multitaper and Conventional Univariate Techniques*, Cambridge Univ. Press, Cambridge, UK, 1993.
- [5] K. S. Riedel and A. Sidorenko, Minimum bias multiple taper spectral estimation, *IEEE Trans. Signal Processing*, Vol. 43, pp. 188-195, 1995.
- [6] D. Slepian, Prolate spheroidal wave functions, Fourier analysis, and uncertainty - V: The discrete case, *Bell Syst. Tech. J.*, Vol. 57, pp. 1371-1430, 1978.
- [7] D. J. Thomson, Spectrum estimation and harmonic analysis, *Proc. IEEE*, Vol. 70, pp. 1055-1096, 1982.
- [8] A. T. Walden, E. J. McCoy, and D. B. Percival, The variance of multitaper spectrum estimates for real Gaussian processes, *IEEE Trans. Signal Processing*, Vol. 42, pp. 479-482, 1995.

# ASYMPTOTIC ANALYSIS OF THE LEAST SQUARES ESTIMATE OF 2-D EXPONENTIALS IN COLORED NOISE

Guy Cohen and Joseph M. Francos

Department of Electrical and Computer Engineering  
Ben-Gurion University  
Beer-Sheva 84105, Israel.

## ABSTRACT

This paper considers the problem of estimating the parameters of complex-valued sinusoidal signals observed in colored noise. This problem is a special case of the general problem of estimating the parameters of a complex-valued homogeneous random field with mixed spectral distribution from a single observed realization of it. The large sample properties of the least squares estimator of the exponentials' parameters are derived, making no assumptions as to the probability distribution of the observed field. It is shown that the least squares estimator is asymptotically unbiased. A simple expression for the estimator asymptotic covariance matrix is derived. The derivation shows that, asymptotically, the least squares estimation of the parameters of each exponential is decoupled from the estimation of the parameters of the other exponentials. Assuming the observed field is a realization of a Gaussian random field, it is further demonstrated that the asymptotic error covariance matrix of the least squares estimate attains the Cramer-Rao bound, even for modest dimensions of the observed field and low signal to noise ratios.

## 1. INTRODUCTION

From the 2-D Wold-like decomposition we have that any 2-D regular and homogeneous discrete random field can be represented as a sum of two mutually orthogonal components: a *purely-indeterministic* field and a *deterministic* one. The purely-indeterministic component has a unique white innovations driven moving average representation. The deterministic component is further orthogonally decomposed into a *harmonic* field

and a countable number of mutually orthogonal *evanescent* fields. In this paper we consider the problem of *least squares* estimation of the parameters of the harmonic component of the field in the presence of the purely-indeterministic component. More specifically, using the results of [2], [3] we evaluate the asymptotic error covariance matrix of the least squares estimator of the harmonic field parameters, from noisy observations of the field. The *colored* observation noise is due to the purely-indeterministic component. This derivation makes no assumptions regarding the probability distribution of the observed field.

## 2. PROBLEM DEFINITION

Let  $\{y(m, n)\}$ ,  $(m, n) \in U$  where  $U = \{(i, j) | 0 \leq i \leq M-1, 0 \leq j \leq N-1\}$  be the observed 2-D complex valued random field such that

$$y(m, n) = h(m, n) + \varepsilon(m, n) \quad (1)$$

and

$$h(m, n) = \sum_{p=1}^P \alpha_p e^{i(\omega_p m + \nu_p n + \varphi_p)} \quad (2)$$

Let  $\theta$  denote the parameter vector of the harmonic field, i.e.,

$$\theta = [\alpha_1 \ \varphi_1 \ \omega_1 \ \nu_1 \ \cdots \ \alpha_P \ \varphi_P \ \omega_P \ \nu_P]^T \quad (3)$$

where  $\alpha_k > 0$ ;  $\varphi_k, \omega_k, \nu_k \in [-\pi, \pi)$ ;  $\omega_k \neq \omega_j$  and  $\nu_p \neq \nu_q$  for  $k \neq j, p \neq q$ .

**Assumption 1:** The purely-indeterministic component  $\{\varepsilon(m, n)\}$  is a circular, zero mean, wide sense homogeneous field, with a positive and piecewise continuous spectral density  $\phi(\omega, \nu)$ , such that the possible discontinuities of  $\phi(\omega, \nu)$  do not coincide with any  $\{(\omega_p, \nu_p)\}_{p=1}^P$ .

**Assumption 2:** The number  $P$  of harmonic components is *a-priori* known.

This work was supported in part by the Israel Ministry of Science under Grant 1233198.



Let  $\mathbf{y}, \mathbf{h}, \boldsymbol{\epsilon}$  denote the observation, harmonic component, and purely-indeterministic component column vectors, respectively, where

$$\mathbf{y} = [y(0,0), \dots, y(M-1,0), y(0,1), \dots, y(M-1,1), \dots, y(0,N-1), \dots, y(M-1,N-1)]^T, \quad (4)$$

and  $\mathbf{h}, \boldsymbol{\epsilon}$  are similarly defined. Let  $\boldsymbol{\Sigma}$  denote the covariance matrix of  $\boldsymbol{\epsilon}$  and hence of  $\mathbf{y}$  as well.

### 3. THE REGRESSION SPECTRUM

Define the  $4P \times 4P$  normalization matrix

$$\mathbf{D}_{M,N} = \text{diag}\{\mathbf{D}, \mathbf{D}, \dots, \mathbf{D}\}, \quad (5)$$

$$\mathbf{D} = \text{diag}\{(MN)^{1/2} (MN)^{1/2} (M^3N)^{1/2} (MN^3)^{1/2}\} \quad (6)$$

Define also the mean gradient vector with respect to the parameter vector  $\boldsymbol{\theta}$

$$\tilde{\boldsymbol{\Phi}}(m,n) = \frac{\partial h(m,n)}{\partial \boldsymbol{\theta}} \quad (7)$$

and let

$$\boldsymbol{\Phi} = \frac{\partial \mathbf{h}}{\partial \boldsymbol{\theta}^T} = \begin{bmatrix} \tilde{\boldsymbol{\Phi}}^T(0,0) \\ \tilde{\boldsymbol{\Phi}}^T(1,0) \\ \vdots \\ \tilde{\boldsymbol{\Phi}}^T(M-1,0) \\ \vdots \\ \tilde{\boldsymbol{\Phi}}^T(M-1,N-1) \end{bmatrix}. \quad (8)$$

Next, consider the sequence of matrices

$$\mathbf{R}_{k,\ell}^{N,M} = \mathbf{D}_{M,N}^{-1} \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \tilde{\boldsymbol{\Phi}}(m+k, n+\ell) \tilde{\boldsymbol{\Phi}}^H(m,n) \mathbf{D}_{M,N}^{-1} \quad (9)$$

Since

$$\lim_{N \rightarrow \infty} \frac{1}{N^{k+1}} \sum_{n=0}^{N-1} n^k e^{i\rho n} = \begin{cases} \frac{1}{k+1}, & \rho = 0 \\ 0, & \rho \neq 0 \end{cases} \quad (10)$$

it can be shown using some straightforward arithmetic that as  $M$  and  $N$  tend to infinity the sequence  $\mathbf{R}_{k,\ell}^{N,M}$  tends to a limit given by

$$\mathbf{R}_{k,\ell} = \text{diag}\{\{e^{i(\omega_p k + \nu_p \ell)} \mathbf{B}_p\}_{p=1}^P\} \quad (11)$$

where

$$\mathbf{B}_p = \begin{bmatrix} 1 & i\alpha_p & \frac{i\alpha_p}{2} & \frac{i\alpha_p}{2} \\ -i\alpha_p & \alpha_p^2 & \frac{\alpha_p^2}{2} & \frac{\alpha_p^2}{2} \\ -\frac{i\alpha_p}{2} & \frac{\alpha_p^2}{2} & \frac{\alpha_p^2}{3} & \frac{\alpha_p^2}{4} \\ -\frac{i\alpha_p}{2} & \frac{\alpha_p^2}{2} & \frac{\alpha_p^2}{4} & \frac{\alpha_p^2}{3} \end{bmatrix} \quad (12)$$

and

$$\begin{aligned} \mathbf{R}_{0,0} &= \lim_{M,N \rightarrow \infty} \mathbf{D}_{M,N}^{-1} \boldsymbol{\Phi}^H \boldsymbol{\Phi} \mathbf{D}_{M,N}^{-1} \\ &= \text{diag}\{\{\mathbf{B}_p\}_{p=1}^P\} \end{aligned} \quad (13)$$

Note that in the terms of [1], [3],  $\mathbf{R}_{k,\ell}$  is a regression correlation matrix.

It can be shown, [3], that  $\mathbf{R}_{k,\ell}$  is a double index positive semi-definite sequence. We therefore conclude using the theorem of Herglotz, Buchner and Weil and following similar arguments to those in [1] p. 45, that  $\mathbf{R}_{k,\ell}$  has a spectral representation of the form

$$\mathbf{R}_{k,\ell} = \frac{1}{4\pi^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} e^{i(k\omega + \ell\nu)} d\mathbf{M}(\omega, \nu) \quad (14)$$

where  $\mathbf{M}(\omega, \nu)$  is a matrix valued function of  $\omega$  and  $\nu$  taking as values Hermitian  $4P \times 4P$  positive semi-definite matrices whose elements are functions of bounded variation, while the functions on the diagonal are non-decreasing. For convenience, we define the regression "spectral density",  $\mathbf{m}(\omega, \nu)$ ,

$$\mathbf{m}(\omega, \nu) = 4\pi^2 \text{diag}\{\{\delta_{\omega_p, \nu_p}(\omega, \nu) \mathbf{B}_p\}_{p=1}^P\} \quad (15)$$

where  $\delta_{\omega_p, \nu_p}(\omega, \nu)$  denotes the Dirac measure concentrated on  $\omega_p, \nu_p$ .

### 4. LEAST SQUARES ESTIMATION OF THE EXPONENTIALS PARAMETERS

Let

$$f_L(\boldsymbol{\theta}) = \frac{1}{2} [\mathbf{y} - \mathbf{h}(\boldsymbol{\theta})]^H [\mathbf{y} - \mathbf{h}(\boldsymbol{\theta})] \quad (16)$$

be the quadratic objective function to be minimized with respect to the parameter vector  $\boldsymbol{\theta}$ .

Assuming  $f_L$  is sufficiently smooth, and employing a second order Taylor series expansion (see [3] for the details) we obtain

$$\hat{\boldsymbol{\theta}} - \boldsymbol{\theta} \cong \mathbf{H}_{\boldsymbol{\theta}}^{-1} \frac{\partial f_L}{\partial \boldsymbol{\theta}} \quad (17)$$

where  $\mathbf{H}_{\boldsymbol{\theta}}$  is the Hessian matrix evaluated at  $\boldsymbol{\theta}$ . Using (17), it is shown in [3] that  $\hat{\boldsymbol{\theta}}$  is an *asymptotically unbiased* estimate of  $\boldsymbol{\theta}$ . The normalized asymptotic error covariance matrix is then given using (17) by

$$\begin{aligned} \text{cov} \hat{\boldsymbol{\theta}} &= \frac{1}{2} [\Re(\mathbf{R}_{0,0})]^{-1} \cdot \\ &\left\{ \lim_{M,N \rightarrow \infty} \mathbf{D}_{M,N}^{-1} \Re(\boldsymbol{\Phi}^H, \boldsymbol{\Phi}) \mathbf{D}_{M,N}^{-1} \right\} [\Re(\mathbf{R}_{0,0})]^{-1} \end{aligned} \quad (18)$$

where we have used the symmetry of  $\Re(\mathbf{R}_{0,0})$ , the existence of its inverse, and the circularity of  $\{\boldsymbol{\epsilon}(m,n)\}$ .

In [3] we show that under the conditions of Assumption 1

$$\lim_{M,N \rightarrow \infty} \mathbf{D}_{M,N}^{-1} \Phi^H \Phi \mathbf{D}_{M,N}^{-1} = \frac{1}{4\pi^2} \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \phi(\omega, \nu) d\bar{\mathbf{M}}(\omega, \nu) \quad (19)$$

Substituting (19) into (18) we have

$$\text{cov } \hat{\boldsymbol{\theta}} = \frac{1}{2} [\Re(\mathbf{R}_{0,0})]^{-1} \frac{1}{4\pi^2} \Re \left[ \int_{-\pi}^{\pi} \int_{-\pi}^{\pi} \phi(\omega, \nu) d\bar{\mathbf{M}}(\omega, \nu) \right] [\Re(\mathbf{R}_{0,0})]^{-1}. \quad (20)$$

Substituting (13) and (15) into (20) we have

$$\text{cov } \hat{\boldsymbol{\theta}} = \text{diag}(\{\mathbf{C}_p\}_{p=1}^P) \quad (21)$$

where

$$\mathbf{C}_p = \frac{1}{2} \begin{bmatrix} \phi(\omega_p, \nu_p) & 0 & 0 & 0 \\ 0 & \frac{7\phi(\omega_p, \nu_p)}{\alpha_p^2} & -\frac{6\phi(\omega_p, \nu_p)}{\alpha_p^2} & -\frac{6\phi(\omega_p, \nu_p)}{\alpha_p^2} \\ 0 & -\frac{6\phi(\omega_p, \nu_p)}{\alpha_p^2} & \frac{12\phi(\omega_p, \nu_p)}{\alpha_p^2} & 0 \\ 0 & -\frac{6\phi(\omega_p, \nu_p)}{\alpha_p^2} & 0 & \frac{12\phi(\omega_p, \nu_p)}{\alpha_p^2} \end{bmatrix} \quad (22)$$

We therefore conclude that, asymptotically, the least squares estimation of the parameters of each exponential is decoupled from the estimation of the parameters of the other exponentials. Moreover, the error variance in estimating the amplitude parameter of each exponential is decoupled and independent of *all* other model parameters. It is a function *only* of the colored noise spectral density at the exponential's frequency. Also, for each exponential the least squares estimation of its two frequency parameters  $\omega_p$  and  $\nu_p$  is asymptotically decoupled. Finally, it should be emphasized that this derivation of the large sample properties of the least squares estimator is independent of the probability distribution function of the observed field.

## 5. ASYMPTOTIC EFFICIENCY OF THE LEAST SQUARES ESTIMATOR

The Cramer-Rao bound (CRB) provides a lower bound on the error variance in estimating the model parameters for any unbiased estimator of these parameters. Since the LS estimator of the harmonic component parameters was shown to be asymptotically unbiased we

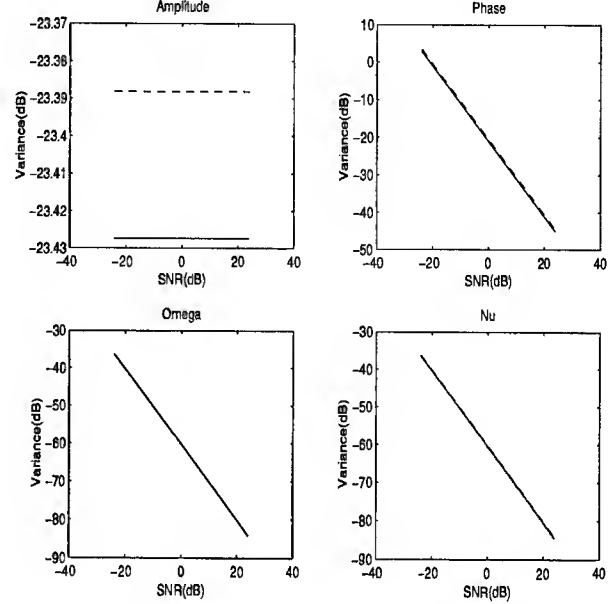


Figure 1: The asymptotic error variance of the LS estimate of the amplitude, phase, and spatial frequency as a function of SNR (dashed line), compared with the corresponding exact CRB (solid line).

investigate in this section its statistical efficiency. Assuming the observed field is Gaussian, we investigate the *asymptotic* performance of the LS estimator of the exponentials' parameters, in comparison with the corresponding exact CRB.

In the first example we investigate the performance as a function of the local signal to noise ratio. The local SNR for the  $k$ th exponential is defined as

$$\text{SNR}_k = 10 \log \frac{\alpha_k^2}{\phi(\omega_k, \nu_k)}. \quad (23)$$

In this example the purely-indeterministic component of the field is a NSHP MA field with support  $S_{1,1}$ . The MA model parameters are  $b(0,1) = -0.9e^{i0.25\pi}$ ,  $b(1,-1) = 0.1e^{i0.4\pi}$ ,  $b(1,0) = -0.5e^{i0.8\pi}$ ,  $b(1,1) = 0.4e^{-i0.2\pi}$ . The driving noise of the MA model is a zero mean, circular, white Gaussian noise field with independent real and imaginary components, each with a unit variance. The harmonic component of the field comprises a single exponential with frequency  $(\omega_1, \nu_1) = (0.6\pi, 0.8\pi)$ . Its amplitude varies to provide the desired range of SNR values. The dimensions of the observed field are  $20 \times 20$ .

The results of this example, Fig. 1, indicate that even for modest dimensions of the observed field, and for a wide range of SNR values, the "asymptotic" error

## 6. CONCLUSIONS

We have investigated the problem of least squares estimation of the parameters of complex-valued exponentials observed in colored noise. Making no assumptions about the probability distribution of the observed field, it is shown that the least squares estimator of the exponentials' parameters is asymptotically unbiased, and a simple expression for its asymptotic covariance matrix is provided. It is further shown that, asymptotically, least squares estimation of the parameters of each exponential is decoupled from the estimation of the parameters of the other exponentials. Moreover, the error variance in estimating the amplitude parameter of each exponential is decoupled and independent of all other model parameters. It is a function only of the colored noise spectral density at the exponential's frequency. Also, for each exponential the least squares estimation of its two frequency parameters  $\omega_p$  and  $\nu_p$  is asymptotically decoupled.

Since the experimental results indicate that even for modest data dimensions the asymptotic covariance matrix of the least squares estimator is very close to the corresponding exact Cramer-Rao lower bound, we conjecture that these results hold when the data dimensions become larger. By definition, at the limit, as data dimensions tend to infinity in both axes, the exact Cramer-Rao bound converges to the asymptotic Cramer-Rao bound. We therefore further conjecture that the asymptotic efficiency of the least squares estimator holds as data dimensions tend to infinity and therefore the asymptotic CRB matrix for the problem of estimating the parameters of 2-D exponentials in colored noise is given by (21)-(22).

## 7. REFERENCES

- [1] U. Grenander and M. Rosenblatt, *Statistical Analysis of Stationary Time Series*, John Wiley & Sons, New York, 1957.
- [2] N. N. Leonenko, "Estimates of linear regression coefficients on a homogeneous random field," *Ukrain. Mat. Z.* vol. 30, pp. 749-756; English transl. in *Ukrainian Math. J.* vol. 30, 1978.
- [3] G. Cohen and J. M. Francos, "Least Squares Estimation of Two-Dimensional Regression Models in Colored Noise," submitted for publication.

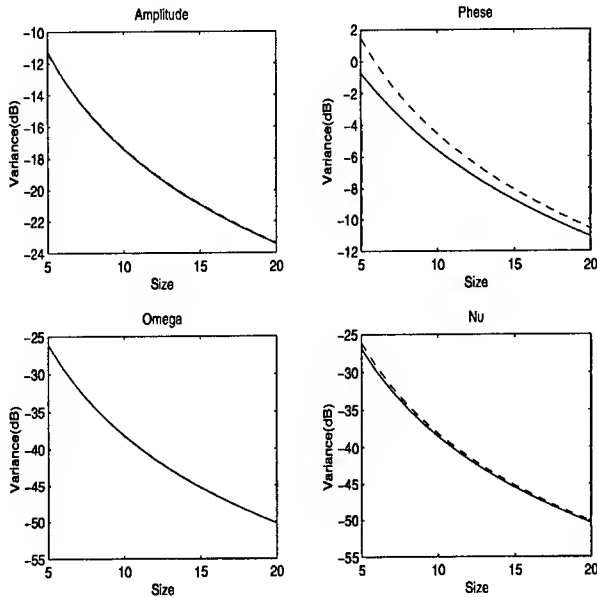


Figure 2: The asymptotic error variance of the LS estimate of the amplitude, phase, and spatial frequency as a function of data dimensions for SNR = -10dB (dashed line), compared with the corresponding exact CRB (solid line).

variances of the LS estimates of the amplitude, phase, and spatial frequency are essentially identical to the corresponding values of the exact CRB. These CRB values are evaluated for the given dimensions of the observed data ( $20 \times 20$  in this case) making no approximations.

In the next example we investigate the effect of the size of the observed field on the performance of the LS estimator and on the CRB. The harmonic component of the field comprises a single exponential such that  $\omega = 0.2\pi$ ,  $\nu = 0.8\pi$ . The purely-indeterministic component is the same as in the first example. To evaluate the functional dependence of the LS estimator asymptotic error variance, and of the corresponding CRB, on the dimensions of the observed field we set  $N = M$  and let both  $N$  and  $M$  assume values from 5 to 20. The results of evaluating the asymptotic error variance of the amplitude, phase, and spatial frequency estimates, and the corresponding CRB, as a function of the field dimensions are depicted in Figure 2 for a local SNR value of -10dB. The results again indicate that the asymptotic error variance of the LS estimator of each of the exponentials' parameters is nearly identical to the corresponding exact CRB, even for modest data dimension and relatively low SNR values.

# Cross-Spectral Methods for Processing Biological Signals

Douglas J. Nelson  
Dept. of Defense  
Ft. Meade, MD 20755, USA  
waveland@erols.com  
www.wavelandplantation.com

**Abstract** ----- We present methods for estimation of signal parameters and apply these methods to biological signals. The methods are based on cross-spectral phase which is computed from the phase of the short time Fourier transform. The methods are applied to acoustical biological signals, including human speech and dolphin sonar clicks. Specifically addressed are the problems of crisp narrow band time frequency representations from very small data sets, accurately estimating speech formants, blind recovery of the group delay of the transmission channel and equalization of time-frequency representations.

**Key Words:** Time-Frequency/Time-Scale analysis, speech, formant recovery, equalization

## (1.0) Introduction

Speech processing has enjoyed a surge in interest in recent years due to an interest within the Government and private industry in the development a machine-based speech processing capability. During this time, speech signal processing research has suffered because the speech research community has accepted MEL-warped cepstral features as the standard signal processing front end and has chosen to focus on language modeling and statistical processing. While these efforts are important, it is the belief of the author that there is still a lot we do not know about speech and other biological signals, and there is a lot of useful information which may be extracted from these signals by appropriate signal processing techniques. This paper is one of a series in which we have attempted to develop new analysis techniques which are effective in parameterizing speech and other non-stationary signals and extracting information contained in the signal itself. In most of these efforts, we have focused on cross-spectral methods based on the phase derivatives of the short time Fourier transform (STFT).

In this paper, we use cross-spectral phase based methods to accurately estimate speech formants, identify and equalize the transmission channel and collapse the excitation function. Since the data were available, the methods are also applied to dolphin clicks and the sonar returns from those clicks. The methods have not been applied to other biological signals, as yet, but because the structures of many biological signals are similar to speech, or at least are consistent with the model used in the analysis presented here, the meth-

ods should apply with equal success. All of the processes presented here are blind, in the sense that the information required to perform the tasks is extracted locally from the signal itself.

## (2.0) The Signal Model

In modeling the speech signal, minimal assumptions are made; however, the model and the processes based on this model are slightly different from the model exploited in standard frame-based speech processing. At least the interpretation of the signal is slightly different.

In normal frame-based processing of speech, the signal is segmented into analysis frames of approximately 25 milliseconds duration. The power spectrum is computed from the windowed signal frames, and the cepstrum is computed from the power spectrum. In the cepstrum computation, the spectrum is effectively smoothed by discarding all but the first few cepstral coefficients (since truncation in time is equivalent to convolution of the spectrum by a *sinc* function.) Implicit in this model is a signal which is the convolution of a nearly periodic function and the impulse response of the vocal tract. The resulting spectrum is the product of the harmonic structure resulting from the excitation function and the spectral response of the vocal tract. In order to recover the frequency response of the vocal tract, the power spectrum is effectively smoothed to remove the harmonic structure of the excitation function. Since phase is discarded, only the magnitude response is estimated. The accuracy, precision, and resolution of the method are all limited by the frequency of the excitation function, the bandwidth of the formants, the length of the analysis window of the Fourier transform, and the smoothing function.

We choose an alternate model of voiced speech in which the vocal tract is excited by a single pulse. In this model, the system is excited as energy from the pulse enters the system. Following the excitation, the system is in nearly steady state resonance damped by the loss of energy in the vocal tract. In this model, speech is essentially a sonar signal in which the formants or resonant frequencies of the vocal tract represent the configuration of the vocal tract, and the ear's function is similar, on a small scale, to that of a bat sounding its environment to fly around objects in the dark, or to a dolphin pulsing an object to identify it by its echoes.

In order to process the signal under this model, an anal-

ysis window which is shorter than the excitation period must be used. Since the excitation frequency  $F_0$  is normally between 100 and 300 Hz, the analysis window must be on the order of a few milliseconds. Without super resolution, the resolution of the power spectrum is no better than few hundred Hertz, and the time resolution of the excitation function is no better than approximately half the length of the analysis window.

For dolphin sonar, the model is nearly identical. The pulse is a short duration single click, which propagates through the water to the target. When the pulse hits the target, the target is excited, followed by a damped resonance.

### (3.0) Cross spectral methods

What we describe now is a method based on the phase gradient which provides a super resolution capability in both time and frequency. Since differentiation in our phase gradient calculation is based on products of short time Fourier transform (STFT) surface, we call the methods cross spectral methods. In the STFT, the Fourier transforms of product of the signal  $f(t)$  and a sequence of time translations of a (short) analysis window  $w(t)$  are computed. The STFT may therefore be represented as

$$F_w(\omega, T) = \int f(t+T)w(-t)e^{-i\omega t} dt. \quad (1)$$

In equation (1), we have followed a convention, which is not quite standard. The order if the surface variables was chosen to represent frequency as column vectors.

We define the channelized instantaneous frequency (CIF) and local group delay (LGD) as

$$\text{CIF}(\omega, T) = \partial_T \arg \{F_w(\omega, T)\} \quad (2)$$

$$\text{LGD}(\omega, T) = -\partial_\omega \arg \{F_w(\omega, T)\} \quad (3)$$

Both the LGD and CIF can be computed as cross spectra

$$\text{CIF}(\omega, T) = \frac{1}{\epsilon} \lim_{\epsilon \rightarrow 0} \arg \left\{ F_w\left(\omega, T + \frac{\epsilon}{2}\right) F_w^*\left(\omega, T - \frac{\epsilon}{2}\right) \right\} \quad (4)$$

$$\text{LGD}(\omega, T) = -\frac{1}{2\pi\epsilon} \lim_{\epsilon \rightarrow 0} \arg \left\{ F_w\left(\omega + \frac{\epsilon}{2}, T\right) F_w^*\left(\omega - \frac{\epsilon}{2}, T\right) \right\} \quad (5)$$

It is easily verified [12] that, for fixed  $\omega_0$ , the STFT  $F_w(\omega, T)$  is the original signal filtered by a filter whose impulse response is

$$w_{\omega_0} = w(t)e^{i\omega_0 t}. \quad (6)$$

i.e.

$$F_w(\omega, T) = f * w_{\omega_0} \quad (7)$$

where "\*" represents convolution. If the filter frequency response

$$W_{\omega_0}(\omega) = \int w_{\omega_0}(t)e^{i(\omega_0 - \omega)t} dt \quad (8)$$

is essentially contained in the positive spectrum i.e.

$$|W_{\omega_0}(\omega)| \gg \max_{\zeta > 0} |W_{\omega_0}(\omega)| \text{ for } \omega < 0, \quad (9)$$

then  $F_w(\omega_0, T)$  is the filtered ANALYTIC representation of the signal, even if the input signal is real [14]. The STFT effectively distributes the analytic signal in time and frequency.

### (4.0) Expected Results

We will now argue that our speech model should result in a process which is the sum of "narrowband processes". One process is a pulse, which is broadband in frequency and relatively localized in time. Following the pulse is a resonance structure, in which there are several resonances or formants, which are normally separated in frequency. During the excitation, the system is driven by the excitation function, and each "filter" of the Fourier transform should resonate at its natural frequency. During the steady state resonance, each filter should respond to the frequency of the resonance which is dominant near that filter frequency. The effects of the channel group delay should be reflected by a relative delay in the filters represented by the STFT. The problem in measuring any of these delays and responses is that they are beyond the resolution of the normal spectrogram. They may, however be estimated by the cross spectral methods.

We assume that at each glottal pulse, the filters respond at their natural frequencies. This means that, at any point on the STFT surface where the signal contribution is dominated by the glottal pulse excitation, the surface may be represented locally as

$$F_{\text{exc}}(\omega, T) = A(\omega, T)e^{i(\omega(T-T_0) - G(\omega) - G_E(\omega))} \quad (10)$$

where  $A(\omega, T)$  is slowly varying in time and frequency,  $T_0$  is the excitation time,

$$g(\omega) = \frac{d}{d\omega} G(\omega) \quad (11)$$

is the channel group delay, and

$$g_E(\omega) = \frac{d}{d\omega} G_E(\omega) \quad (12)$$

is the group delay of the vocal tract at excitation. In this case, the LGD and CIF are

$$\text{LGD}_{\text{exc}}(\omega, T) = T_0 - T + g(\omega) + g_E(\omega) \quad (13)$$

$$\text{CIF}_{\text{exc}}(\omega, T) = \omega \quad (14)$$

where  $T_0$  is the time of occurrence of the pulse. Note that this means that we should expect the LGD to collapse the pulse to a single curve

$$T(\omega) = T_0 - g(\omega) - g_E(\omega), \quad (15)$$

where the right hand side is not dependent on time.

Now consider a region on the STFT surface near a formant in steady state resonance. In this case, there is no contribution from the glottal pulse, and all filters near the formant will be pulled to the formant resonant frequency. In this case, the STFT surface may be locally modeled as

$$F_{res}(\omega, T) = A(\omega, T) e^{i(\omega_0(T-T_0) - G(\omega) - G_R(\omega))}, \quad (16)$$

where the group delay of the vocal tract at resonance  $g_R$  is the derivative of  $G_R(\omega)$ . The LGD and CIF are

$$LGD_{res}(\omega, T) = g(\omega) + g_R(\omega) \quad (17)$$

$$CIF_{res}(\omega, T) = \omega_0. \quad (18)$$

#### (4.1) Mixed Partial

The mixed partial derivatives give us a convenient test for excitation and steady state resonance. In the excitation case, we may compute the expected mixed phase partial derivatives as

$$E\{\partial_\omega \partial_T F_{exc}(\omega, T)\} = E\{\partial_T \partial_\omega F_{exc}(\omega, T)\} = 1. \quad (19)$$

In the steady state resonance case, we may compute the expected mixed phase partial derivatives as

$$E\{\partial_\omega \partial_T F_{res}(\omega, T)\} = E\{\partial_T \partial_\omega F_{res}(\omega, T)\} = 0 \quad (20)$$

With these two relationships, we may build indicator functions to test whether any point on the STFT surface is the response to an excitation pulse or a resonance of the vocal tract. The functions

$$I_E(\omega, T) = |1 - \partial_T \partial_\omega F_w(\omega, T)| \quad (21)$$

$$I_R(\omega, T) = |\partial_T \partial_\omega F_w(\omega, T)| \quad (22)$$

have expected values zero at excitation and resonance, respectively. If the condition is not met at the point  $(\omega_0, T_0)$ , then the STFT response at that point is not driven by excitation (resonance), and is therefore driven by another process, such as resonance (excitation), or noise or interference. We can therefore discard the points on the STFT surface which are not indicated as the signal condition we seeking, and therefore improve the processing gain. An important observation is that the indicator functions serve to effectively partition the STFT surface into three surfaces. On one surface, the excitation is dominant, and resonance is effectively removed. On the second surface, resonance is dominant, and excitation is effectively removed. And, on the third surface, artifacts other than resonance and excitation are dominant.

To test the indicator functions, the two indicator functions were computed and compared to the remapped STFT

surface. An example of an indicator surfaces is represented by Figures 5. As can be seen, the surfaces do indeed indicate the excitation and resonance correctly, and the indicators tend to be mutually exclusive. That is, the excitation indicator tends to reject resonance, and the resonance indicator tends to reject the excitation.

As an additional check, the phase of the mixed partial surface was plotted in a neighborhood of both the excitation and resonances. The displays clearly show that the mixed partials behave as predicted by the model.

#### (4.2) Equalization

Finally, we address the problem of blind recovery of the channel group delay. The ability to blindly equalize a channel containing a biological signal was first discovered while analyzing dolphin clicks. Note that if the channel group delay were zero, then the STFT surface in a neighborhood of excitation and steady state resonance respectively would be

$$F_{res}(\omega, T) = A(\omega, T) e^{i\omega(T-T_0)} \quad (23)$$

$$F_{exc}(\omega, T) = A(\omega, T) e^{i\omega_0(T-T_0)} \quad (24)$$

The respective local group delays would be

$$LGD_{exc}(\omega, T) = T_0 - T \quad (25)$$

$$LGD_{res}(\omega, T) = 0. \quad (26)$$

We start by estimating the group delay of the vocal tract. To do this, we have ground truth in the form of TIMIT data which was collected under studio recording conditions, in which we may assume that the channel effects are insignificant. Several portions of voiced speech from the TIMIT database were processed by remapping the surface to correct the LGD and CIF. In each case, the formants collapsed to constant frequency "lines", and the excitation pulses collapsed to broadband impulses in time, or constant time "lines". This established that the group delay of the vocal tract is zero.

What we therefore must do is calculate an estimated group delay function, which effectively forces the collapsed excitation pulses on the equalized TF surface to impulses in time. Recall that the clean TIMIT data satisfied this condition. For small group delays, where the channel group delay is less than the analysis window, we can effectively equalize by identifying, with the aid of the indicator function, a  $T_0$  near the center of the excitation pulse. If we consider the surface

$$F_{eq}(\omega, T) = e^{-i \arg F_w^*(\omega, T)} F_w(\omega, T), \quad (27)$$

we see that conditions (25,26) are satisfied. That is, correcting the spectral phase by the observed spectral phase of a glottal pulse effectively removes the group delay, at least for group delays which are relatively well behaved. The process was used to "straighten out" NTIMIT (Figures 8 and 9) and dolphin backscatter data. The process has not been tested on severe channels, since no data were available.

For larger group delays, it is necessary to piece together

several spectra to reconstruct the group delay. If the analysis window is less than the span of the channel group delay, the estimated group delay will have an ambiguity. A delay greater than the length of the analysis window will result in the group delay aliasing or being folded modulo the length of the analysis window. In addition, the LGD function is only valid where there is significant spectral energy near the formants. In the nulls between the formants, the energy is low, resulting in erroneous estimation. Both of these ambiguities can be resolved by combining estimates from several spectra.

### (5.0) Methods and Conclusions

Following the discussion in the paper, samples of data from the TIMIT and NTIMIT database were selected. The TIMIT data was recorded with a close talking microphone, and the NTIMIT data is the TIMIT data subjected to the NYNEX telephone channel. In addition, random samples of dolphin back scatter data were selected from data provided by the Department of the Navy. The STFT, LGD, CIF, and mixed partial surfaces were computed, with a prolate spheroidal window of the same approximate length as the expected excitation pulse. The surfaces were remapped as

$$F_{\text{remap}}(\text{CIF}(\omega, T), T + \text{LGD}(\omega, T)) = F_w(\omega, T) \quad (28)$$

The excitation pulses of the remapped NTIMIT surfaces collapsed to curves. The dolphin excitation pulses were slightly curved, and the TIMIT pulses were nearly straight. Equalization of the pulses resulted in nearly straight lines after remapping. In each case, the formants collapsed to nearly constant frequency lines. Typical examples of the processed data are represented in the figures.

The second partial derivatives were computed for selected samples of TIMIT data, and the data depicted in the figures verify the mixed partial relationships described above.

### Bibliography

- [1] G. von Békésy, *Experiments in Hearing*, Mc Graw Hill, 1960.
- [2] L. Cohen, *Time-Frequency Analysis*, Prentice-Hall, (Englewood Cliffs, NJ), 1995.
- [3] Helmholtz, *On the Sensations of Tone*, 1863.
- [4] H. Fletcher, *Speech and Hearing in Communication*, 2nd ed., Van Nostrand, 1953
- [5] Rayleigh, Lord, "On our perception of sound direction," *Philosophical Mag.*, 13, pp. 214-232, 1907.
- [6] S. Stevens and J. Volkman, "The relation of pitch to frequency," *Amer. Journal of Psych.*, vol. 53, pp. 329, 1940.
- [7] D. Nelson, "Special Purpose Correlation Functions for Improved Signal Detection and Parameter Estimation," *Proc. IEEE Conf. on Acoust., Speech and Sig. Proc.*, pp.73-76, April, 1993.
- [8] D.J. Nelson, "Correlation-based Formant Recovery," *Proc. IEEE Conf. Acoust., Speech and Sig. Proc.*, 1997.
- [9] D. Nelson and J. Pencak, "Pitch based methods for speech analysis," *Proc SPIE Adv Sig Proc Conf*, San Diego, CA. vol. 2303, 1994.
- [10] D.J. Nelson, "Estimation of FM Modulation of Multi-Component Signals from the Fourier Phase" *Proc. IEEE Conf. Acoust, Speech and Sig Proc.* vol. 6, pp. 3421-3424, 1998.
- [11] D.J. Nelson and O.P. Kenny, "Invertible Time-Frequency Representations", *Proc. SPIE Adv. Sig. Proc. Conf.*, July, 1998.
- [12] D.J. Nelson, "Invertible Time-Frequency Surfaces", *Proc. IEEE Conf. on Time-Frequency and Time-Scale*, Pittsburgh, October, 1998.
- [13] D.J. Nelson and W. Wysocki, "Cross-spectral methods with an application to speech processing", *Proc. SPIE Adv. Sig. Proc. Conf.*, July, 1999.
- [14] D. Nelson, G. Cristobal, V. Kober, F. Cakrak, P. Loughlin and L. Cohen, "Denoising Using Time-Frequency and Image Processing Methods", *Proc. of the SPIE Adv. Sig. Proc. Conf.*, July, 1999.

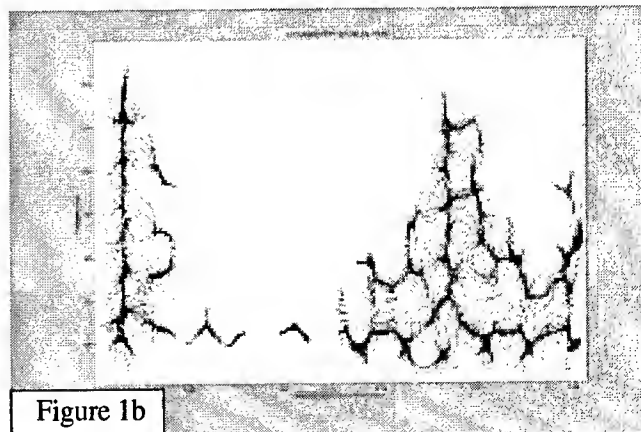
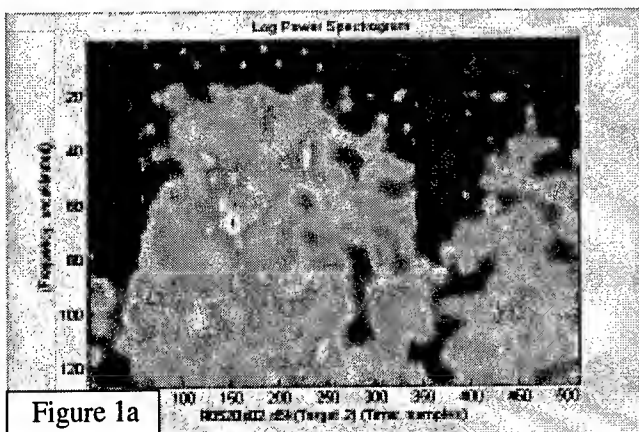


Figure 1a: Conventional spectrogram of dolphin click and sonar return.

Figure 1b: Focused and equalized STFT computed from 512 data samples.



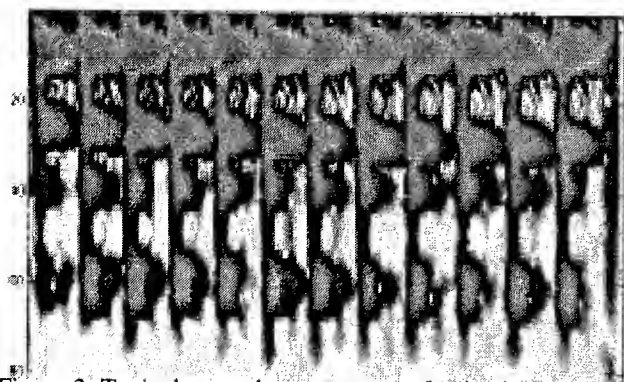


Figure 2: Typical normal spectrogram of voiced speech.

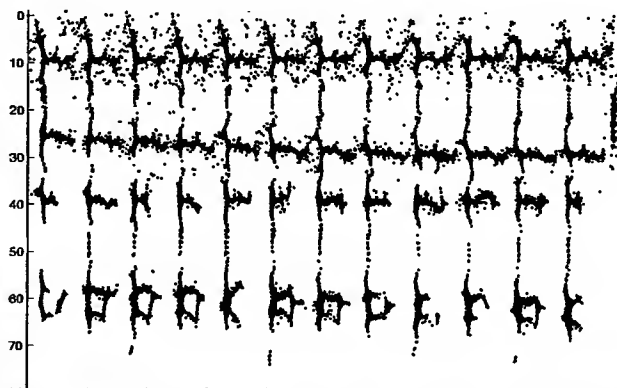


Figure 3: Typical voiced speech focused spectrogram showing collapsed glottal pulses and focused formants.

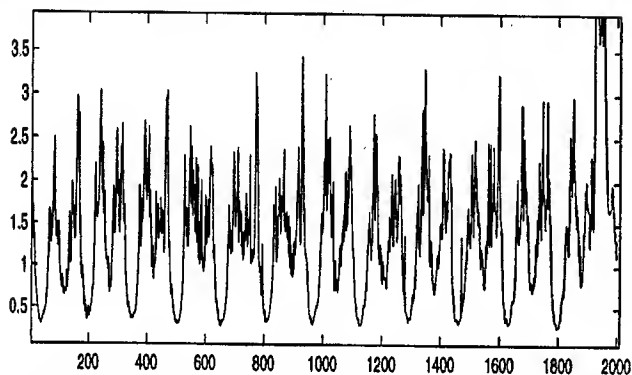


Figure 4: Mixed phase partial derivatives averaged over frequency. Nulls indicate glottal pulses

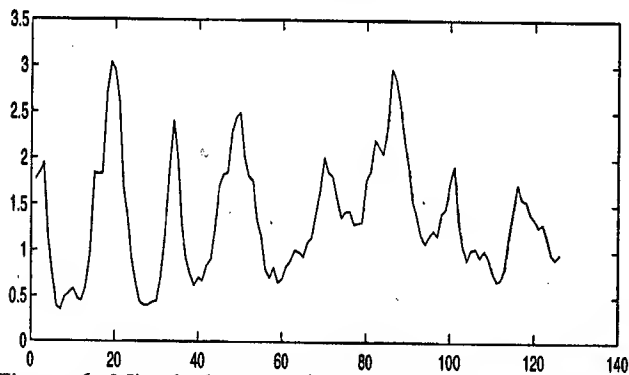


Figure 6: Mixed phase partial derivatives averaged over time. Nulls indicate formant energy bands.

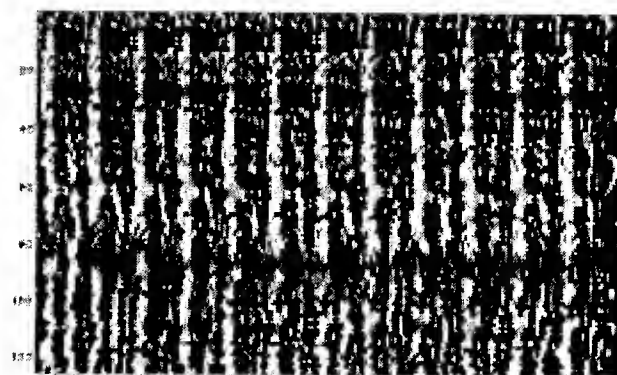


Figure 5: Un-remapped mixed phase partial derivative surface of equalized speech. Channel may be estimated in white area.

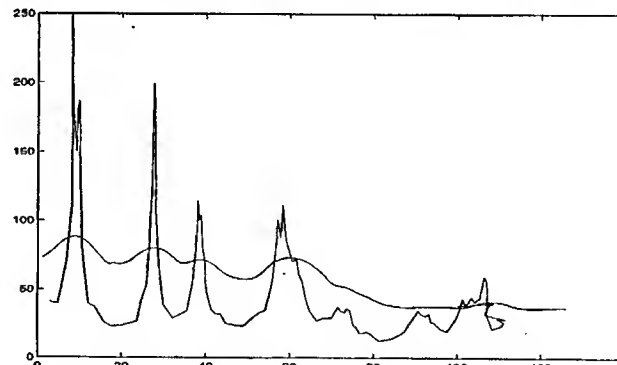


Figure 7: Power spectrum (dB) and remapped averaged spectrum (dB), weighted by mixed partials

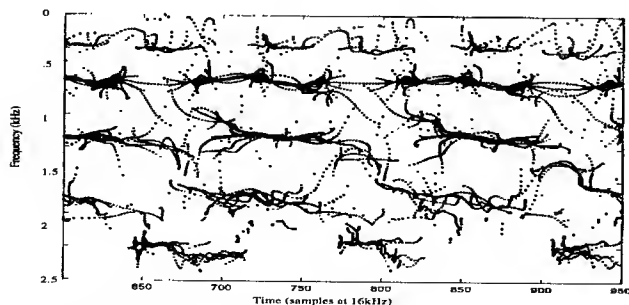


Figure 8: Un-equalized NTIMIT data

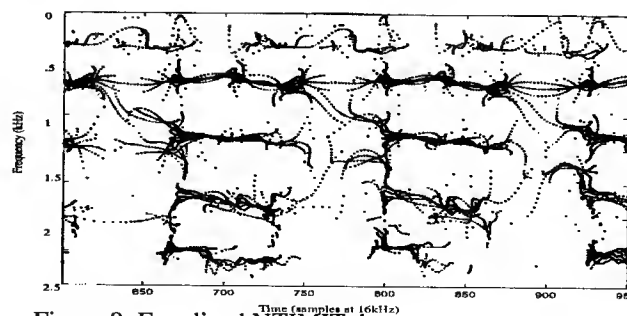


Figure 9: Equalized NTIMIT data



# DEFAULT PRIOR FOR ROBUST BAYESIAN MODEL SELECTION OF SINUSOIDS IN GAUSSIAN NOISE

Christophe Andrieu<sup>†</sup> - J.-M. Pérez<sup>‡</sup>

<sup>†</sup>Signal Processing Group, Engineering Dept. Cambridge University  
Trumpington Street, CB2 1PZ Cambridge, UK.

<sup>‡</sup>Dept. Cómputo Científico y Estadística and Centro de Estadística y Software Matemático  
(CESMA) Universidad Simón Bolívar. Aptdo. 89000 Caracas 1080A, Venezuela.

Email: ca226@eng.cam.ac.uk - jperez@cesma.usb.ve

## ABSTRACT

We address the problem of detection and estimation of sinusoids embedded in white Gaussian noise. We follow a Bayesian approach and adopt robust default priors, Expected Posterior priors. In order to compute the associated Bayes factor required for model selection we resort to Monte Carlo Markov chain algorithms, and illustrate performance on an example.

## 1 Introduction

Model selection is a fundamental data analysis task. It has many applications in various fields of science and engineering, including the canonical problem of detection and estimation of sinusoids embedded in noise. Over the past two decades, many of the classical model selection problems have been addressed using information criteria such as AIC [3], BIC [18] or Rissanen's MDL [15]. The widespread use of these criteria is mainly due to their intrinsic simplicity. However they rely on asymptotic expansions, when the number of data is large, and quantifying the effect of these approximations when small data set are analyzed seems to be difficult. Bayesian statistics provides a simple and sound framework to the task of model selection, see [5] for a recent review. Unfortunately, within this framework, model selection appears more difficult from a practical point of view, mainly for two reasons. Firstly, Bayesian inference requires the choice of a prior distribution for the unknown parameters, which might at first sight appear to be a difficult exercise, especially in situations where no prior information is available. Many efforts have been devoted to the development of a methodology that provides a framework for the automatic determination of such uninformative prior, see [12] for a review. However, most of these priors are typically improper<sup>1</sup>, which does not cause any problem when parameter estimation is concerned, but can lead to indeterminate answers when model selection is investigated, as illustrated in Section 3. Secondly it is worth noting that the quantities required to perform Bayesian model selection do not usually admit any closed-form expression and that analytical approximations, such as BIC [18], or numerical evaluations are then required.

In this paper we propose to address the problem of robust and consistent detection of the number of sinusoids embedded in noise

in a Bayesian framework using uninformative improper priors. A review of the literature on the subject can be found in [1]. Our approach relies on expected posterior (EP) priors that have been recently introduced in [13] and numerical techniques, Monte Carlo Markov chains (MCMC), that have revolutionized applied statistics over the past ten years. The problem considered is of great interest in many fields, as suggested by the vast literature dedicated to the problem (see for example [7], [8], [14] and references therein), but it should be pointed out that the methodology can be adapted to other scenarios.

The paper is organized as follows. In Section 2 the signal model is given. In Section 3, we specify robust and uninformative prior distributions for our problem. In Section 4 we develop MCMC algorithms to compute the quantities required to perform Bayesian model selection. The performance of our procedure is illustrated by computer simulations in Section 5.

## 2 Model of the data

Let  $\mathbf{y} = (y_1, y_2, \dots, y_T)^T$  be an observed vector of  $T$  real data samples. The elements of  $\mathbf{y}$  may be represented by different models  $\mathcal{M}_k$  corresponding either to samples of noise only or to the superimposition of  $k$  sinusoids corrupted by noise:

$$\begin{aligned} \mathcal{M}_0 : y_t &= n_{t,0} \\ \mathcal{M}_k : y_t &= \sum_{j=1}^k (a_{c_{j,k}} \cos[\omega_{j,k}t] + a_{s_{j,k}} \sin[\omega_{j,k}t]) + n_{t,k}, \end{aligned}$$

where  $\omega_{j_1,k} \neq \omega_{j_2,k}$  for  $j_1 \neq j_2$  and  $a_{j,k}$ ,  $\omega_{j,k}$  are respectively the amplitude and the radial frequency of the  $j^{\text{th}}$  sinusoid for the model with  $k$  sinusoids. The noise sequence  $\mathbf{n}_k \triangleq (n_{1,k}, \dots, n_{T,k})$  is assumed zero-mean white Gaussian with covariance matrix  $\sigma_k^2 \mathbf{I}_T$ . In a vector-matrix form, we have

$$\mathbf{y} = \mathbf{D}(\boldsymbol{\omega}_k) \mathbf{a}_k + \mathbf{n}_k,$$

where  $[\mathbf{a}_k]_{2i-1,1} \triangleq a_{c_{i,k}}$ ,  $[\mathbf{a}_k]_{2i,1} \triangleq a_{s_{i,k}}$  and  $[\boldsymbol{\omega}_k]_{i,1} \triangleq \omega_{i,k}$  for  $i = 1, \dots, k$ . The matrix  $\mathbf{D}(\boldsymbol{\omega}_k)$  is defined as  $[\mathbf{D}(\boldsymbol{\omega}_k)]_{t,2j-1} = \cos[\omega_{j,k}t]$  and  $[\mathbf{D}(\boldsymbol{\omega}_k)]_{t,2j} = \sin[\omega_{j,k}t]$  for  $t = 1, \dots, T$ ,  $j = 1, \dots, k$ . This allows us to write the likelihood of the observations

$$p(\mathbf{y} | \mathbf{a}_k, \sigma_k^2, \boldsymbol{\omega}_k) = \frac{1}{(2\pi\sigma_k^2)^{T/2}} \exp\left(-\frac{\|\mathbf{y} - \mathbf{D}(\boldsymbol{\omega}_k)\mathbf{a}_k\|_2^2}{2\sigma_k^2}\right).$$

We assume here that the number  $k$  of sinusoids and their parameters  $\boldsymbol{\theta}_k \triangleq (\mathbf{a}_k, \sigma_k^2, \boldsymbol{\omega}_k)^T$  are unknown. Given the data set  $\mathbf{y}$ , our

C. Andrieu is sponsored by AT&T Laboratories, Cambridge UK.

<sup>1</sup>A distribution is improper when its sum is not finite. As a consequence such a distribution cannot be normalized.

objective is to estimate  $k$  and  $\theta_k$ . We will assume that the maximum possible number of sinusoids is  $k_{\max} \triangleq \lfloor (T-1)/2 \rfloor$ , see [1] for motivations.

### 3 Model selection and EP priors

We follow a Bayesian approach where the unknowns  $k$  and  $\theta_k$  are regarded as being a priori distributed according to appropriate prior distributions. These priors reflect our degree of belief of the relevant values of the parameters. In this section we first recall the key role played by Bayes factors for model selection and point out the problem associated with the use of improper priors. Then EP priors are introduced to solve our problem. This section ends with the detection objectives for the problem investigated, formulated in a Bayesian framework.

#### 3.1 Bayesian model selection and EP priors

Assume as in our case that  $k_{\max}$  models  $\mathcal{M}_k$  are under consideration for a data set  $\mathbf{y}$ . Model  $\mathcal{M}_k$  corresponds to assuming a probability density  $p(\mathbf{y}|\theta_k)$  for the observations, the likelihood, which depends on a parameter  $\theta_k$ . The Bayesian approach requires the specification of prior densities  $p_k(\theta_k) \triangleq p(\theta_k|k)$  and possibly the specification of model prior probabilities  $p(k)$ . Then the key quantities on which Bayes model selection relies are the Bayes factors

$$B_{ij} = \frac{p(i|\mathbf{y})/p(j)}{p(j|\mathbf{y})/p(i)},$$

which by introduction of the *predictive densities*

$$m_i(\mathbf{y}) = \int_{\Theta_i} p(\mathbf{y}|\theta_i) p_i(\theta_i) d\theta_i,$$

can be reformulated as

$$B_{ij} = m_i(\mathbf{y}) / m_j(\mathbf{y}),$$

which shows that specification of the model prior probability is not necessary. The Bayes factor is often interpreted as the “odds provided by the data for  $\mathcal{M}_i$  versus  $\mathcal{M}_j$ .” Motivations for the choice of Bayes factors for model selection can be found in [5], [11]. The cornerstone of this approach seems at first sight to be the choice of a prior, especially in situations when no prior knowledge is available. The need for automatic or *default* approaches for the choice of uninformative prior has been recognized for a long time [10]. In estimation problems, the use of vague or uninformative prior distributions, including sometimes improper prior distributions, is typically a satisfactory solution [12]. When performing model selection, however, one has to be much more careful as default priors are typically improper, and, thus, depend on arbitrary multiplicative constants, i.e.  $p_i^N(\theta_i) = c_i f_i^N(\theta_i)$  for some function  $f_i^N$  (we use the superscript  $N$  to indicate the use of a *uninformative* or *default prior* for the model parameters). Hence, the resultant Bayes factor

$$B_{ij}^N = \frac{c_i}{c_j} \frac{\int_{\Theta_i} p(\mathbf{y}|\theta_i) p_i^N(\theta_i) d\theta_i}{\int_{\Theta_j} p(\mathbf{y}|\theta_j) p_j^N(\theta_j) d\theta_j}, \quad (1)$$

is indeterminate, and cannot be used for model selection. Note that the use of “vague proper priors” usually give wrong answers in Bayesian model selection, as it has long been recognized since [10]. A number of proposals to overcome this problem have been made. Approaches using conventional priors have been studied in [10], [20]. In the case of nested models, proper hierarchical robust prior models have been successfully developed for various applications [1], [16]. Other approaches include the Intrinsic Bayes Factor (IBF) [4], the Fractional Bayes Factor (FBF) and

the method suggested in [17], among others. Most of these later methods deal with the problem by rescaling the Bayes factor by a *correction factor* in such a way that any undesirable constant cancels. EP priors have been recently proposed in [13] and belong to this family of approach. This approach relies on the utilization of the device of “imaginary training sample”, a well known approach [9]. More precisely imagine that some extra data  $\mathbf{y}^*$  are available, and that these data are used to update a possibly non proper prior  $p_i^N(\theta_i) = c_i f(\theta_i)$  using Bayes rule,

$$p^N(\theta_i|\mathbf{y}^*) = \frac{p(\mathbf{y}^*|\theta_i) c_i f(\theta_i)}{\int_{\Theta_i} p(\mathbf{y}^*|\theta_i) c_i f(\theta_i) d\theta_i}.$$

This posterior distribution, when it is defined, does not depend on  $c_i$  anymore and can be used as a new prior distribution. However as  $\mathbf{y}^*$  is not actually observed these posteriors are not available. The key idea of EP prior is to consider a suitable *predictive measure* on the imaginary training sample space  $\mathcal{Y}^*$  and integrate out these artificial data, leading to

$$p^*(\theta_i) = \int_{\mathcal{Y}^*} p^N(\theta_i|\mathbf{y}^*) m^*(d\mathbf{y}^*).$$

The measure  $m^*$  can intuitively be viewed as arising from beliefs of how a real training set would behave. Several choices for  $m^*$  are possible, but we note that  $\mathbf{y}^*$  and  $m^*$  should be such that  $p^N(\theta_i|\mathbf{y}^*)$  exists. In many cases for example it is required that the size of  $\mathbf{y}^*$  is more than a given *minimal size* so that the posterior  $p^N(\theta_i|\mathbf{y}^*)$  exists. A training set with such size is said to be of *minimal size*. We now detail two possibilities retained in this paper.

An attractive choice for  $\mathbf{y}^*$  and  $m^*$  consists of selecting a *base model*  $\mathcal{M}_*$  and defining  $m^*(\mathbf{y}^*) \triangleq \int_{\Theta} p(\mathbf{y}^*|\theta) p_*(\theta) d\theta$ . Intuitively the base model should be at least as simple as the other models, so that little constraint on  $\mathbf{y}^*$  is imposed. In the case of nested models, the EP priors resulting from this choice correspond to the intrinsic priors for the Arithmetic IBF [13]. Alternatively an empirical version of  $m^*$  can also be considered, where training samples are obtained by resampling from the observations  $\mathbf{y}$ . Once proper  $\mathbf{y}^*$  and  $m^*$  have been selected, the Bayes factor of  $\mathcal{M}_i$  against  $\mathcal{M}_j$  resulting from the EP priors can be expressed as

$$B_{ij}^*(\mathbf{y}) = \frac{m_{p_i^*}(\mathbf{y})}{m_{p_j^*}(\mathbf{y})}, \text{ with } m_{p_i^*}(\mathbf{y}) \triangleq \int_{\Theta_i} p(\mathbf{y}|\theta_i) p_i^*(\theta_i) d\theta_i,$$

Therefore, the resulting Bayes factor does not depend on arbitrary multiplicative constants leading to consistent Bayesian model selection. We list here some of the other interesting properties of EP priors:

- The resulting Bayesian inference allows for multiple comparisons. For instance  $B_{ij}$  will be equal to  $B_{ik}$  times  $B_{kj}$ , a property not shared by all default model selection methods.
- In many cases, it is possible to find  $m^*$  such that, for a sample of *minimal size*, there is *predictive matching* for the comparisons of model  $\mathcal{M}_i$  against  $\mathcal{M}_j$ , i.e., the Bayes factor  $B_{ij}$  is equal to 1.
- In certain situations, the approach is essentially equivalent to previous successful approaches. For instance in the case of nested models, when  $\mathcal{M}_1$  is nested in every other model, choosing  $m^*$  to be the marginal of  $\mathbf{y}^*$  under  $\mathcal{M}_1$  is asymptotically equivalent to the arithmetic IBF [4].

Other nice properties of EP priors can be found in [13]. We now derive an EP prior for the problem of detection of sinusoids in noise.

### 3.2 EP prior for robust spectral analysis

For  $\theta_k$ , we assume the following uninformative prior distribution for all the dimensions

$$p^N(\mathbf{a}_k, \sigma_k^2, \omega_k | k) \propto \frac{1}{\sigma_k^2} \mathbb{I}_{\Omega}(k, \omega_k),$$

which is clearly improper. The set  $\Omega$  is defined as  $\bigcup_{k=0}^{k_{\max}} \{k\} \times \Omega_k$  where  $\Omega_k \triangleq \{\omega_k \in (0, \pi)^k; i \neq j \text{ implies } [\omega_k]_i \neq [\omega_k]_j\}$  for  $k > 0$  and  $\Omega_0 \triangleq \emptyset$ . We do not want to favor any subset of the artificial data set  $\mathbf{y}^*$  and thus introduce the set  $\mathcal{Y}_k^*$  of all subvectors of length  $m_k$  ( $m_k$  will be determined later on) made from  $\mathbf{y}^*$ . In order to define all the quantities related to each elements of  $\mathcal{Y}_k^*$  we introduce an arbitrary labelling on all the combinations of length  $m_k$  in the set  $\{1, \dots, m_{k_{\max}}\}$ . The vector  $\mathbf{y}_{k,l}^*$  is then the vector of length  $k$  made from  $\mathbf{y}^*$  for which the retained indices correspond to combination number  $l$ . From now on  $l$  is assumed to be random, distributed according to a uniform distribution. We now first compute the expression of  $p(\mathbf{a}_k, \sigma_k^2, \omega_k | k, \mathbf{y}_{k,l}^*)$  obtained from Bayes' rule. After some algebra one obtains for  $k > 0$

$$p(\sigma_k^2 | k, \omega_k, \mathbf{y}_{k,l}^*) = \frac{\left(\frac{\mathbf{y}_{k,l}^{*T} \mathbf{P}_{k,l}^* \mathbf{y}_{k,l}^*}{2}\right)^{\frac{m_k-2k}{2}} \exp\left[\frac{-\mathbf{y}_{k,l}^{*T} \mathbf{P}_{k,l}^* \mathbf{y}_{k,l}^*}{2\sigma_k^2}\right]}{\Gamma((m_k-2k)/2) (\sigma_k^2)^{(m_k-2k)/2+1}} \times \frac{\exp\left[\frac{-\|\mathbf{a}_k - \mathbf{m}_{k,l}^*\|^2}{2\sigma_k^2}\right]}{|2\pi\sigma_k^2 \mathbf{M}_{k,l}^*|^{1/2}},$$

$$p(\mathbf{a}_k | k, \sigma_k^2, \omega_k, \mathbf{y}_{k,l}^*) = \frac{1}{|2\pi\sigma_k^2 \mathbf{M}_{k,l}^*|^{1/2}},$$

$$p(\omega_k | \mathbf{y}_{k,l}^*) = 1/\pi^k,$$

when  $\mathbf{y}_{k,l}^* \notin \text{span}\{\mathbf{D}(\omega_k)\}$  and  $m_k - 2k \geq 0$ , and where

$$\mathbf{M}_{k,l}^{*-1} = \mathbf{D}_{k,l}^{*T}(\omega_k) \mathbf{D}_{k,l}^*(\omega_k), \mathbf{m}_{k,l}^* = \mathbf{M}_{k,l}^* \mathbf{D}_{k,l}^{*T}(\omega_k) \mathbf{y}_{k,l}^*,$$

$$\mathbf{P}_{k,l}^* = \mathbf{I}_{m_k} - \mathbf{D}_{k,l}^*(\omega_k) \mathbf{M}_{k,l}^{*-1} \mathbf{D}_{k,l}^{*T}(\omega_k),$$

with  $\mathbf{D}_{k,l}^*(\omega_k)$  the  $m_{k_{\max}} \times 2k$  matrix extracted from  $[\mathbf{D}(\omega_k)]_{1:m_{k_{\max}}, 1:2k}$  corresponding to the  $l^{\text{th}}$  combination of indices. When  $k = 0$  then

$$p(\sigma_0^2 | 0, \mathbf{y}_{0,l}^*) = \frac{\left(\frac{\mathbf{y}_{0,l}^{*T} \mathbf{y}_{0,l}^*}{2}\right)^{m_0/2} \exp\left[\frac{-\mathbf{y}_{0,l}^{*T} \mathbf{y}_{0,l}^*}{2\sigma_0^2}\right]}{\Gamma(m_0/2) (\sigma_0^2)^{m_0/2+1}}.$$

Two possibilities for the definition of  $m^*$  can be proposed:

- Choose  $\mathcal{M}_0$  as base model, and build the measure

$$m^*(\mathbf{y}^*) = \int_{\mathbb{R}^+} p(\mathbf{y}^* | \sigma_0^2) \frac{d\sigma_0^2}{\sigma_0^2} = \frac{\Gamma(m_{k_{\max}}/2)}{(\pi \mathbf{y}^{*T} \mathbf{y}^*)^{m_{k_{\max}}/2}}.$$

One observes here that  $m^*(\mathbf{y}^*)$  is spread out, reflecting the fact that little constraints are imposed by the base model. We can now give the definition of the EP prior

$$p^*(\mathbf{a}_k, \sigma_k^2, \omega_k | k) = \sum_{l=1}^{\#\mathcal{Y}_k^*} \int_{\mathbb{R}^{m_{k_{\max}}}} p(\mathbf{a}_k, \sigma_k^2, \omega_k | k, \mathbf{y}_{k,l}^*) p(l) m^*(d\mathbf{y}^*),$$

and we observe that  $m_k = 2k$  is the minimum size required on  $\mathbf{y}_{k,l}^*$  for  $p(\mathbf{a}_k, \sigma_k^2, \omega_k | k, \mathbf{y}_{k,l}^*)$  to exist. Note that this prior is not proper, but does not introduce any arbitrary constant in the Bayes factors.

- Build  $m^*(\mathbf{y}^*)$  from the observations, i.e.

$$m^*(\mathbf{y}^*) = \frac{1}{\#\mathcal{Y}_{k_{\max}}} \sum_{\mathbf{y}_l \in \mathcal{Y}_{k_{\max}}} \mathbb{I}_{\{\mathbf{y}_l\}}(\mathbf{y}^*).$$

We will here only explore the first possibility.

### 3.3 Model order prior

The model order prior distribution does not affect the Bayes factors, and thus the model selection rule, but can have an influence on the mixing properties of the algorithm developed later. We chose here a truncated Poisson distribution, i.e.  $p(k) \propto \frac{1}{k!} \mathbb{I}_{\{0, \dots, k_{\max}\}}$  but any other choice such as a uniform distribution would have been possible.

### 3.4 Integration of the nuisance parameters

The proposed Bayesian model allows for the integration of the so-called nuisance parameters,  $\mathbf{a}_k$  and  $\sigma_k^2$ , and subsequently to obtain an expression for  $p(k, \omega_k, l, \mathbf{y}^* | \mathbf{y})$  up to a normalizing constant. According to Bayes' theorem

$$p(k, \mathbf{a}_k, \sigma_k^2, \omega_k, l, \mathbf{y}^* | \mathbf{y}) \propto p(\mathbf{y} | k, \mathbf{a}_k, \sigma_k^2, \omega_k) \times p(\mathbf{a}_k, \sigma_k^2, \omega_k | k, \mathbf{y}_{k,l}^*) m(\mathbf{y}^*) p(k) p(l),$$

with

$$\mathbf{M}_{k,l}^{-1} = \mathbf{D}^T(\omega_k) \mathbf{D}(\omega_k) + \mathbf{M}_{k,l}^{*-1}$$

$$\mathbf{m}_{k,l} = \mathbf{M}_{k,l} [\mathbf{D}^T(\omega_k) \mathbf{y} + \mathbf{M}_{k,l}^{*-1} \mathbf{m}_{k,l}^*].$$

Consequently,

$$p(k, \mathbf{a}_k, \sigma_k^2, \omega_k, l, \mathbf{y}^* | \mathbf{y}) \propto \frac{\exp\left[\frac{-(\mathbf{a}_k - \mathbf{m}_{k,l})^T \mathbf{M}_{k,l}^{-1} (\mathbf{a}_k - \mathbf{m}_{k,l})}{2\sigma_k^2}\right]}{|2\pi\sigma_k^2 \mathbf{M}_{k,l}^{-1}|^{1/2} (2\pi\sigma_k^2)^{T/2}} \times \frac{\exp\left[\frac{-\frac{1}{2}(\mathbf{y}^T \mathbf{y} - \mathbf{m}_{k,l}^T \mathbf{M}_{k,l}^{-1} \mathbf{m}_{k,l} + \mathbf{y}_{k,l}^{*T} \mathbf{y}_{k,l}^*)}{2\sigma_k^2}\right]}{\Gamma((m_k-2k)/2) (\sigma_k^2)^{(m_k-2k)/2+1}} \times \left(\frac{\mathbf{y}_{k,l}^{*T} \mathbf{P}_{k,l}^* \mathbf{y}_{k,l}^*}{2}\right)^{(m_k-2k)/2} p(k) p(l) m(\mathbf{y}^*) \frac{1}{\pi^k} \mathbb{I}_{\Omega}(k, \omega_k).$$

The integration of  $\mathbf{a}_k$  (similar to a normal distribution) and then of  $\sigma_k^2$  (similar to an inverse gamma distribution) yields for  $k \geq 1$

$$p(k, \omega_k, l, \mathbf{y}^* | \mathbf{y}) \propto |\mathbf{M}_{k,l}^*|^{-1/2} |\mathbf{M}_{k,l}|^{1/2} \frac{\Gamma((T+m_k-2k)/2)}{\Gamma((m_k-2k)/2)} \times \left(\frac{\mathbf{y}_{k,l}^{*T} \mathbf{P}_{k,l}^* \mathbf{y}_{k,l}^*}{2}\right)^{(m_k-2k)/2} p(k) p(l) m(\mathbf{y}^*) \frac{1}{\pi^k} \mathbb{I}_{\Omega}(k, \omega_k) \times \left(\frac{\mathbf{y}^T \mathbf{y} - \mathbf{m}_{k,l}^T \mathbf{M}_{k,l}^{-1} \mathbf{m}_{k,l} + \mathbf{y}_{k,l}^{*T} \mathbf{y}_{k,l}^*}{2}\right)^{-\frac{T+m_k-2k}{2}}. \quad (2)$$

and a similar expression for  $k = 0, [2]$ . The overall parameter space  $\Theta$  can be written as a countable union of subspaces  $\Theta = \bigcup_{k=0}^{k_{\max}} \{k\} \times \Theta_k$  where  $\Theta_0 \triangleq \mathbb{R}^+ \times \mathcal{Y}^*$  for  $k = 0$ ,  $\Theta_k \triangleq (\mathbb{R}^2)^k \times \mathbb{R}^+ \times \Omega_k \times \mathcal{C}_k \times \mathcal{Y}^*$ , for  $k \in \{1, \dots, k_{\max}\}$  and where  $\mathcal{C}_k = \{1, \dots, C_{m_{k_{\max}}}^k\}$ .

### 3.5 Estimation Objectives and Bayesian computation

The objective is to compute the Bayes factors, and more precisely the quantities  $m_{p_i}^*(\mathbf{y})$ . Then model selection can be performed from these quantities, see Section 5 for example.

### 4 Bayesian computation

In order to evaluate the Bayes factors, we are interested in computing the quantities  $p(k | \mathbf{y})$  for which no closed-form expression exists. One has to resort to numerical methods. MCMC techniques are very powerful methods that allow for these quantities to be computed in an efficient manner. Roughly speaking MCMC

consist of running an ergodic Markov chain whose invariant distribution is the distribution of interest, here the posterior distribution  $p(k, \mathbf{a}_k, \sigma_k^2, \omega_k, l, \mathbf{y}^* | \mathbf{y})$ . Under weak conditions the sample path of the generated Markov can be used to compute quantities related to the posterior distribution. In our case, for example, we are interested in the marginal posterior distribution  $p(k = j | \mathbf{y})$ , which can be evaluated from the sample path of the Markov chain using the formula  $\hat{p}_i(k = j | \mathbf{y}) = \frac{1}{i - i_0 + 1} \sum_{l=i_0}^i \mathbb{I}_{\{j\}}(k^{(l)})$ , after convergence towards the invariant distribution. The algorithm we develop here is an adaptation of the algorithm presented in [1], that takes here into account the specific features introduced by the use of EP priors. For a complete introduction to MCMC one should however refer for example to [16] and references therein.

#### 4.1 The main algorithm

In order to build the Markov chain for our problem we introduce the following updates of the parameters: (a) birth of a new sinusoid, (b) death of an existing sinusoid, (c) update of the frequencies for all the sinusoids one-at-a-time, when  $k \neq 0$  (d) Update the training samples  $\mathbf{y}^*$ . The birth and death moves perform dimension changes respectively from  $k$  to  $k + 1$  and  $k$  to  $k - 1$ . These moves are defined by heuristic considerations, the only condition to be fulfilled being to maintain the correct invariant distribution. A particular choice will only have influence on the convergence rate of the algorithm. Other moves may be proposed, but we have found that the ones suggested here lead to satisfactory results. The resulting transition kernel of the simulated Markov chain is then a mixture of the different transition kernels associated with the moves described above. This means that at each iteration one of the candidate moves: birth, death or update is randomly chosen. The probabilities for choosing these moves are  $b_k$ ,  $d_k$  and  $u_k$  respectively, such that  $b_k + d_k + u_k = 1$  for all  $0 \leq k \leq k_{\max}$  and the update moves are performed at each iteration. The move is performed if the algorithm accepts it. For  $k = 0$  the death move is impossible, so that  $d_0 \triangleq 0$ . For  $k = k_{\max}$  the birth move is impossible and thus  $b_{k_{\max}} \triangleq 0$ . Except in the cases described above, we take the following probabilities:

$$b_k \triangleq c \min \left\{ 1, \frac{p(k+1)}{p(k)} \right\}, d_{k+1} \triangleq c \min \left\{ 1, \frac{p(k)}{p(k+1)} \right\},$$

where  $p(k)$  is the prior probability of model  $\mathcal{M}_k$  and  $c$  is a parameter which tunes the proportion of dimension/update move. The algorithm can be summarized as follows:

---

#### Reversible Jump MCMC algorithm

---

1. Initialization: set  $(k^{(0)}, \theta_k^{(0)}) \in \Theta$ .
2. Iteration  $i$ :  
Choose one of the following move
  - With probability  $b_{k(i)}$  a "birth" move (See Subsection 4.5).
  - With probability  $d_{k(i)}$  a "death" move (See Subsection 4.5).
  - With probability  $u_{k(i)}$  update the frequencies  $\omega_k$  (See Subsection 4.2)

Update  $l, \mathbf{y}^*$  (See Subsection 4.4). ■

We describe more precisely these different moves below. In what follows, in order to simplify notation, we drop the superscript  $^{(i)}$  from all variables at iteration  $i$ .

#### 4.2 Updating the frequencies

We use the same technique as described in [1] with the target distribution here proportional to (2) in order to take into account the EP prior.

#### 4.3 Updating the nuisance parameters

In this subsection we show how it is possible to sample the nuisance parameters. We point out that if one is not interested in estimating these nuisance parameters then this simulation step is not required. We obtain by straightforward calculations:

$$\begin{aligned} \sigma_k^2 | (\mathbf{y}, k, \omega_k, \mathbf{y}_{k,l}^*) &\sim \mathcal{IG} \left( \frac{T+m_k-2k}{2}, \frac{\mathbf{y}^T \mathbf{y} - \mathbf{m}_{k,l}^T \mathbf{M}_{k,l}^{-1} \mathbf{m}_{k,l} + \mathbf{y}_{k,l}^{*T} \mathbf{I}_{m_k} \mathbf{y}_{k,l}^*}{2} \right) \\ \mathbf{a}_k | (\mathbf{y}, k, \sigma_k^2, \omega_k, \mathbf{y}_{k,l}^*) &\sim \mathcal{N}(\mathbf{m}_{k,l}, \sigma_k^2 \mathbf{M}_{k,l}). \end{aligned}$$

#### 4.4 Update the $\mathbf{y}^*$

---

##### Update $l, \mathbf{y}^*$

---

- Draw  $\tilde{\sigma}_k^2 | (\mathbf{y}, k, \omega_k) \sim \mathcal{IG} \left( \frac{T-2k}{2}, \frac{\mathbf{y}^T \mathbf{y} - \mathbf{y}^T \mathbf{D}(\omega_k) [\mathbf{D}^T(\omega_k) \mathbf{D}(\omega_k)]^{-1} \mathbf{D}^T(\omega_k) \mathbf{y}}{2} \right)$ ,  
 $\tilde{\mathbf{a}}_k | (\mathbf{y}, k, \sigma_k^2, \omega_k) \sim \mathcal{N} \left( [\mathbf{D}^T(\omega_k) \mathbf{D}(\omega_k)]^{-1} \mathbf{D}^T(\omega_k) \mathbf{y}, \sigma_k^2 [\mathbf{D}^T(\omega_k) \mathbf{D}(\omega_k)]^{-1} \right)$ .
- Draw  $\tilde{l} \sim \mathcal{U}_{\# \mathcal{Y}_k}$ .
- Draw  $\tilde{\mathbf{y}}_{k,l}^* \sim \mathcal{N}(\mathbf{D}_{k,l}^*(\omega_k) \tilde{\mathbf{a}}_k, \tilde{\sigma}_k^2 \mathbf{I}_{m_k})$ .
- Random walk:  $(\tilde{\mathbf{y}}_{k,l}^*)^c \sim \mathcal{N}((\mathbf{y}_{k,l}^*)^c, \lambda \mathbf{I}_{m_{k_{\max}} - m_k})$ .
- Accept update with probability  $\min \{1, r_{\mathbf{y}^*}\}$ , see [2]. ■

where  $\lambda$  is user defined.

#### 4.5 Reversible jumps: birth/death moves

Suppose that the current state of the Markov chain is in  $\{k\} \times \Theta_k$ , then

---

##### Birth move

---

- Propose a new frequency at random on  $(0, \pi)$ :  $\omega \sim \mathcal{U}_{(0, \pi)}$ .
- Propose new values for  $\mathbf{y}_{k+1,l}^*$  with the strategy of Subsection 4.4.
- Evaluate  $\alpha_{birth}$ , see [2].
- Accept  $(k+1, \omega_{k+1})$ , with probability  $\alpha_{birth,k}$  else remain at  $(k, \omega_k)$ . ■

Assume that the current state of the Markov chain is in  $\{k+1\} \times \Theta_{k+1}$ , then

---

##### Death move

---

- Choose a sinusoid at random among the  $k+1$  existing sinusoids:  $j \sim \mathcal{U}_{\{1, \dots, k+1\}}$ .
  - Propose new values for  $\mathbf{y}_{k-1,l}^*$  with the strategy of Subsection 4.4.
  - Evaluate  $\alpha_{death}$ , see [2].
  - Accept  $(k, \omega_k)$  with probability  $\alpha_{death,k}$  else remain  $(k+1, \omega_{k+1})$ . ■
-

## 5 Simulations

A set of synthetic data was generated. The data set corresponds to a sinusoid series with four frequencies and same amplitudes, contaminated with Gaussian noise. The SNR was 0dB. The MCMC algorithm described above was run for 50000 iterations, of which, the first 5000 were burnt out. The maximum number of frequencies was set at  $k_{\max} = 8$ , and the mean parameter for the truncated Poisson was set at  $\Lambda = 4$ . The renormalized quantities  $p(k|y)/p(k)$  that allow us to compute the Bayes factors are shown in Tab. 1. In Fig. 1 we present for each iteration the estimate of  $p(k|y)/p(k)$  for  $k = 0, \dots, 8$ . Comparison with the results obtained in [1] with slightly informative priors are currently investigated.

0	1	2	3	4	5 $\geq$
3.91%	6.54%	13.92%	29.34%	35.97%	10.32%

Table 1: Renormalized predictives  $p(k|y)/p(k)$ .

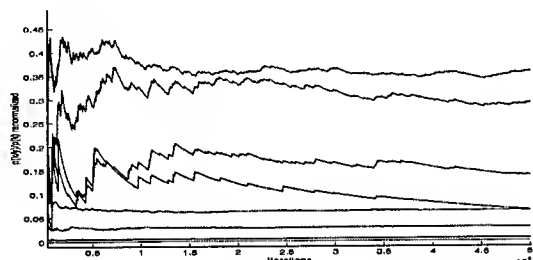


Figure 1: For the iterations  $i=1, \dots, 50000$ , the current estimates of  $p(k|y)/p(k)$

## 6 REFERENCES

- [1] C. Andrieu and A. Doucet, "Joint Bayesian model selection and estimation of noisy sinusoids via reversible jump MCMC", *IEEE Trans. Sig. Proc.*, vol. 47, no. 10, 2667-2676, 1999.
- [2] C. Andrieu and J. M. Pérez, "Default Prior for Robust Bayesian Model Selection of Sinusoids in Gaussian Noise," *Tech. Rep. CUED*, Cambridge 2000.
- [3] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Auto. Cont.*, vol. 19, no. 6, pp. 716-723, 1974.
- [4] J.O. Berger and L.R. Pericchi, "The intrinsic Bayes factor for model selection and prediction", *J. Am. Stat. Assoc.*, 91, 109-122, 1996.
- [5] J. Berger, and L. Pericchi, "Objective Bayesian methods for model selection: introduction and comparison," *ISDS Discussion Paper 00-09*.
- [6] J.M. Bernardo, "Reference prior distributions for Bayesian inference", *J. Roy. Stat. Soc. B*, 41, 113-147, 1979.

- [7] G.L. Bretthorst, *Bayesian Spectrum Analysis and Parameter Estimation*, Lecture Note in Statistics, vol. 48, Springer-Verlag, New-York, 1988.
- [8] P.M. Djurić, "A model selection rule for sinusoids in white Gaussian noise," *IEEE Trans. Sig. Proc.*, vol. 44, no. 7, pp. 1744-1751, 1996.
- [9] I.J. Good, *Probability and the weighting of evidence*, Haffner New-York, 1950.
- [10] H. Jeffreys, *Theory of Probability*, Oxford University Press, 1961.
- [11] R.E. Kass and A.E. Raftery, "Bayes factors," *J. Am. Stat. Assoc.*, 90, 773-796, 1995.
- [12] R.E. Kass and L. Wasserman, "The Selection of prior distributions by formal rules", *J. Am. Stat. Assoc.*, 91, 1343-1369, 1996.
- [13] J.M. Pérez and J. Berger, "Expected posterior prior distributions for model selection," *ISDS Discussion Paper 00-08*, 2000.
- [14] D.C. Rife and R.R. Boorstyn, "Multiple-tone parameter estimation from discrete-time observations," *Bell Syst. Tech. J.*, vol 55, pp. 1389-1410, 1976.
- [15] J. Rissanen, "Modeling by shortest data description," *Automatica*, vol. 14, 1978, pp. 465-471.
- [16] C.P. Robert and G. Casella, *Monte Carlo Statistical Methods*, Springer-Verlag, 1999.
- [17] D.J. Spiegelhalter and A.F.M. Smith, "Bayes factor for linear and for log-linear models with vague prior information", *J. Roy. Stat. Soc. B*, 44, 377-387, 1982.
- [18] G. Schwartz, "Estimating the Dimension of a Model", *Ann. Stat.* 6, pp. 461-464, 1978.
- [19] L. Tierney, "Markov chains for exploring posterior distributions," *Ann. Stat.*, pp. 1701-1728, 1994.
- [20] A. Zellner and A. Siow, "Posterior odds ratios for selected regression hypotheses", in *Bayesian Statistics*, J.M. Bernardo, M.H. DeGroot, D.V. Lindley and A.F.M. Smith (eds.).

# ON THE EXACT SOLUTION TO THE “GLIDING TONE” PROBLEM

*Lorenzo Galleani and Leon Cohen*

City University of New York, 695 Park Ave., New York, NY 10021 USA.

## ABSTRACT

The gliding tone problem is the response of a resonant circuit when the driving force is a gliding tone, that is a chirp. The problem was first considered by Barber and Ursell, and independently by Hok, both papers appearing in 1948. Barber and Ursell, and Hok, and subsequent investigators considered approximate solutions and attempted to qualitatively understand the behavior of the response. An exact solution has never been obtained. We have found the *exact* Wigner distribution of the solution. This allows one to study the nature of the solution without any approximations. We have obtained the exact solution by way of a new method to study dynamical systems.

## 1. INTRODUCTION

In 1948 Barber and Ursell [1] and independently Hok [4] considered the problem of the response of a harmonic oscillator to a “gliding tone”.<sup>1</sup> Specifically the issue is the behavior of the solution to a resonant circuit [1, 4, 5]

$$\frac{d^2x(t)}{dt^2} + 2\mu \frac{dx(t)}{dt} + \omega_0^2 x(t) = f(t) \quad (1)$$

with

$$f(t) = e^{j\beta t^2/2} \quad (2)$$

Subsequent to Barber and Ursell, and Hok, many investigators have considered this problem in a variety of contexts and have tried to qualitatively understand the solution and also obtain approximate solutions. An exact solution to this problem has not been achieved. We, also, have not been able to obtain an exact explicit solution; but we have been able to obtain the exact solution to the Wigner distribution of  $x(t)$ !

Galleani's permanent address: Dipartimento di Elettronica, Politecnico di Torino, C.so Duca degli Abruzzi 24, 10129 Torino, Italy.

Work supported by the Office of Naval Research, the NASA JOVE, and the NSA HBCU/MI programs.

<sup>1</sup>The phrase “gliding tone” was used by Barber and Ursell.

We have been able to obtain the exact solution by using a new method that we have developed to study dynamical systems [3].

In the next section we give the explicit solution to the gliding tone problem and subsequently we give a few numerical examples. In the appendix we explain how we have obtained the exact solution.

## 2. THE EXACT WIGNER DISTRIBUTION

We define the Wigner distribution by [2]

$$W(t, \omega) = \frac{1}{2\pi} \int x^*(t - \frac{1}{2}\tau) x(t + \frac{1}{2}\tau) e^{-j\tau\omega} d\tau \quad (3)$$

and the step function,  $u(t)$ , by

$$u(t) = \begin{cases} 1 & t \geq 0 \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

We now give the exact solution of the Wigner distribution of  $x(t)$  which satisfies Eq. (20). The proof is outlined in the appendix. Explicitly,

$$W(t, \omega) = \frac{2}{|\beta|} \frac{u(\tau)}{z_2 - z_1} \left[ \frac{1}{\bar{z}_1 - z_1} \left( \frac{e^{-2z_1\tau} - e^{-2\bar{z}_2\tau}}{\bar{z}_2 - z_1} - \frac{e^{-2\bar{z}_1\tau} - e^{-2\bar{z}_2\tau}}{\bar{z}_2 - \bar{z}_1} \right) - \frac{1}{\bar{z}_1 - z_2} \left( \frac{e^{-2z_2\tau} - e^{-2\bar{z}_2\tau}}{\bar{z}_2 - z_2} - \frac{e^{-2\bar{z}_1\tau} - e^{-2\bar{z}_2\tau}}{\bar{z}_2 - \bar{z}_1} \right) \right] \quad (5)$$

with

$$\tau = t - \omega/\beta \quad (6)$$

and

$$z_1 = -j\omega + \mu - \sqrt{\mu^2 - \omega_0^2} \quad (7)$$

$$\bar{z}_1 = j\omega + \mu - \sqrt{\mu^2 - \omega_0^2} \quad (8)$$

$$z_2 = -j\omega + \mu + \sqrt{\mu^2 - \omega_0^2} \quad (9)$$

$$\bar{z}_2 = j\omega + \mu + \sqrt{\mu^2 - \omega_0^2} \quad (10)$$

### 3. UNDERDAMPED, OVERDAMPED, AND CRITICALLY DAMPED CASES

We now explicitly specialize to the underdamped, overdamped, and critically damped cases. As is standard we define the critical frequency,  $\omega_c$ , for these three cases

$$\begin{aligned}\omega_c &= \sqrt{\omega_0^2 - \mu^2} & \mu < \omega_0 & \text{Underdamped} \\ \omega_c &= \sqrt{\mu^2 - \omega_0^2} & \mu > \omega_0 & \text{Overdamped} \\ \omega_c &= 0 & \mu = \omega_0 & \text{Critically damped}\end{aligned}$$

The explicit Wigner distributions are  
*Underdamped:*

$$W(t, \omega) = \frac{1}{2|\beta|\omega_c} u(\tau) e^{-2\mu\tau} \times \left[ \frac{\sin(2(\omega - \omega_c)\tau)}{\omega(\omega - \omega_c)} - \frac{\sin(2(\omega + \omega_c)\tau)}{\omega(\omega + \omega_c)} \right] \quad (11)$$

*Overdamped:*

$$W(t, \omega) = \frac{1}{|\beta|} u(\tau) e^{-2\mu\tau} \times \left[ \frac{\sin(2\omega\tau) \cosh(2\omega_c\tau)}{\omega(\omega^2 + \omega_c^2)} - \frac{\cos(2\omega\tau) \sinh(2\omega_c\tau)}{\omega_c(\omega^2 + \omega_c^2)} \right] \quad (12)$$

*Critically Damped:*

$$W(t, \omega) = \frac{1}{|\beta|} u(\tau) e^{-2\mu\tau} \frac{\sin(2\omega\tau) - 2\omega\tau \cos(2\omega\tau)}{\omega^3} \quad (13)$$

In the above solutions there are singularities at some values of  $\omega$ . We give the limits at those singular values for the three cases:

*Underdamped:*

$$\lim_{\omega \rightarrow \pm\omega_c} W(t, \omega) = \frac{1}{2|\beta|\omega_c} u(\tau) e^{-2\mu\tau} \left[ \frac{4\omega_c\tau - \sin(4\omega_c\tau)}{2\omega_c^2} \right] \quad (14)$$

$$\lim_{\omega \rightarrow 0} W(t, \omega) = \frac{1}{|\beta|} u(t) e^{-2\mu t} \left[ \frac{\sin(2\omega_c t) - 2\omega_c t \cos(2\omega_c t)}{\omega_c^3} \right] \quad (15)$$

*Overdamped :*

$$\lim_{\omega \rightarrow 0} W(t, \omega) = \frac{1}{|\beta|} u(t) e^{-2\mu t} \times$$

$$\left[ \frac{2\omega_c t \cosh(2\omega_c t) - \sinh(2\omega_c t)}{\omega_c^3} \right] \quad (16)$$

*Critically Damped:*

$$\lim_{\omega \rightarrow 0} W(t, \omega) = \frac{8}{3} \frac{1}{|\beta|} u(t) t^3 e^{-2\mu t} \quad (17)$$

### 4. EXAMPLES

We now give some examples which indicates in broad terms the behavior of the solution. In a subsequent publication the nature of the solutions will be studied in detail. For all the examples we take  $\beta = 1$  and  $\omega_0 = 18$ . We then vary the "damping" coefficient  $\mu$  to study the three cases typical of the harmonic oscillator.

#### 4.1. Underdamped Case

We compute the Wigner in Eq. (11) choosing  $\mu = 1$ . The results are plotted in Fig. 1. Several important observations can be made. First the dashed line represents the instantaneous frequency of the forcing chirp, that is  $\omega_i(t) = \beta t$ . The chirp is concentrated only along this line, because its representation in the Wigner distribution domain is  $\delta(\omega - \beta t)$ . The gray scale image is the actual Wigner distribution of  $x(t)$ . For every fixed  $\omega$  one can notice that the Wigner distribution starts always after the chirp. Hence the Wigner distribution reproduces the causal behavior of the harmonic oscillator, that is, the system gives no output until there is no input.

This feature can be easily understood also from (11), where the step function  $u(t - \omega/\beta)$  guarantees this causal behavior. The response of the system is mainly concentrated around the critical frequency  $\omega_c$ , while it's weaker at all the other frequencies. This happens because the classic transfer function of the harmonic oscillator has a peak at  $\omega = \omega_c$ , and the energy of the chirp, that is almost constant over the entire frequency axis is amplified at the critical frequency. The transfer function goes to zero for  $\omega \rightarrow \infty$ , and that is why the Wigner distribution goes to zero for  $\omega \gg \omega_c$ . Also, observing the limit (14) at  $\omega = \omega_c$ , one can see that the Wigner distribution has an exponential damping factor, where the damping coefficient is  $2\mu$ ; which is twice the damping of the free oscillation of the system. Near  $\omega = 0$  the Wigner distribution presents the characteristic cross terms. In this particular case, they are generated by the interference of the energy concentrations at  $\omega = \pm\omega_c$ . (In the plots only  $\omega \geq 0$  is shown, since the Wigner distribution is symmetrical about  $\omega = 0$ .)

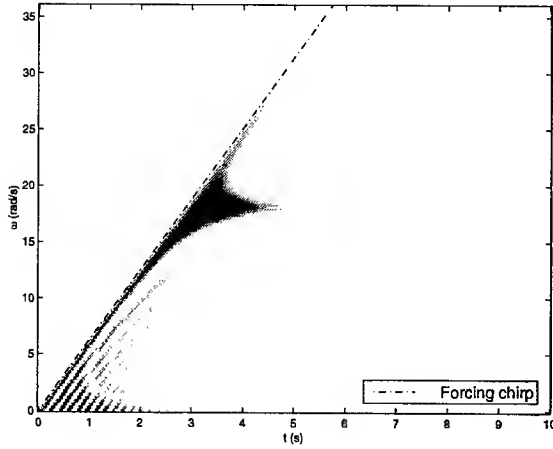


Figure 1: Underdamped case:  $\mu = 1$ . The Wigner distribution in (11) is plotted taking  $\omega_0 = 18$  and  $\beta = 1$ . The gray scale image is the actual Wigner, while the dashed line is the instantaneous frequency  $\omega_i(t) = \beta t$  of the input chirp  $f(t)$ . Note the energy concentration around the resonant frequency  $\omega_c = \sqrt{\omega_0^2 - \mu^2} \approx \omega_0$ . The oscillating terms around  $\omega = 0$  are the cross terms generated by the symmetric energy concentrations at  $\omega = \pm\omega_c$  (only  $\omega \geq 0$  is shown).

#### 4.2. Critically and Overdamped Case

In Fig. 2 and 3 we plot the Wigner distribution for  $\mu = 18$  and  $\mu = 30$ , respectively for the critically and overdamped cases. The two images are very similar, and some of the remarks made for the underdamped case are still valid here. The response is again causal, but no energy concentration is present at the critical frequency values. This is in accordance with the anharmonic behavior of the oscillator for these damping values which is well known [4]. Also the cross terms near  $\omega = 0$  disappear, due to the lack of the energy distributions. Again the amplitude of the Wigner distribution varies in relation to the modulus of the transfer function of the system.

### 5. CONCLUSION

We believe that we have effectively achieved the exact solution to the gliding tone problem. Our method presents a new perspective to studying dynamical problems. That is, instead of directly seeking the solution for the dynamical variable, we seek directly the Wigner distribution of the variable. Remarkably, sometimes that is easier than finding the solution itself. The gliding tone problem is such a case.

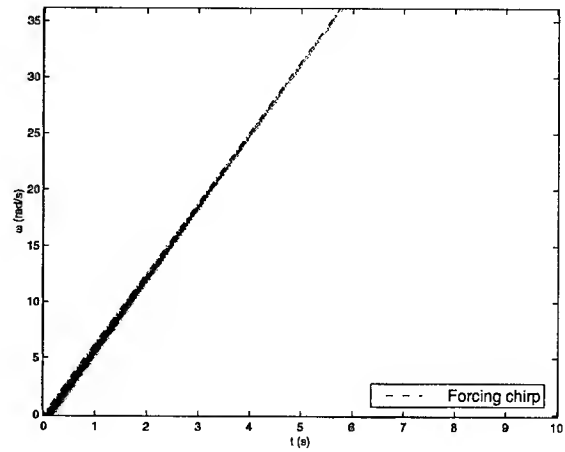


Figure 2: Critically damped case:  $\mu = 18$ . We plot the analytic Wigner distribution in (13), with  $\omega_0$  and  $\beta$  as in Fig. 1. The energy concentrations clearly visible in the underdamped case have disappeared, in accordance with the anharmonic behavior of the oscillator in the critical damping case. Also the cross terms have disappeared.

### 6. APPENDIX

Using the notation

$$D = \frac{d}{dt} \quad (18)$$

we rewrite Eq. (1) as

$$[D^2 + 2\mu D + \omega_0^2]x(t) = f(t) \quad (19)$$

and then we factorize the differential operator acting on  $x(t)$

$$[D - p_1][D - p_2]x(t) = f(t)$$

where

$$p_{1,2} = -\mu \pm \sqrt{\mu^2 - \omega_0^2}$$

The associated Wigner distribution equation to the problem (19) is [3]

$$[A^2 + 2\mu A + \omega_0^2][B^2 + 2\mu B + \omega_0^2]W_{x,x}(t, \omega) = W_{f,f}(t, \omega) \quad (20)$$

where

$$W_{f,f} = \delta(\omega - \beta t) = \frac{1}{|\beta|} \delta(t - \omega/\beta)$$

The equation for the Wigner distribution can be factorized as

$$[A - p_1][A - p_2][B - p_1][B - p_2]W_{x,x}(t, \omega) = W_{f,f}(t, \omega) \quad (21)$$



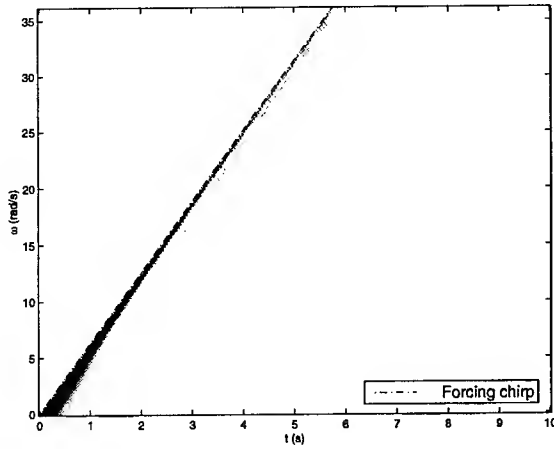


Figure 3: Overdamped case with  $\mu = 30$ . We plot the Wigner distribution in (12), with  $\omega_0$  and  $\beta$  as in Fig. 1. Similar considerations to Fig. 2 hold, and in particular the energy is mainly concentrated around  $\omega = 0$ .

The solution is obtained by first transforming the equation in the equivalent system

$$\begin{aligned} [A - p_1] W_1 &= W_{f,f} \\ [A - p_2] W_2 &= W_1 \\ [B - p_1] W_3 &= W_2 \\ [B - p_2] W_{x,x} &= W_3 \end{aligned}$$

Substituting for  $A$  and  $B$  and collecting the terms we have

$$\begin{aligned} \frac{\partial W_1}{\partial t} + 2z_1 W_1 &= 2W_{f,f} \\ \frac{\partial W_2}{\partial t} + 2z_2 W_2 &= 2W_1 \\ \frac{\partial W_3}{\partial t} + 2\bar{z}_1 W_3 &= 2W_2 \\ \frac{\partial W_{x,x}}{\partial t} + 2\bar{z}_2 W_{x,x} &= 2W_3 \end{aligned}$$

where the coefficients  $z_1, z_2, \bar{z}_1, \bar{z}_2$  are defined by Eqs. (7). The equations are considered as ordinary differential equations, and solved setting the constant of integration to zero. In a forthcoming publication we will prove that this approach is equivalent to the Wigner distribution of the impulse response method [6].

As an example we give the derivation of  $W_1$  from the first equation

$$W_1 = e^{-2z_1 t} \int_{-\infty}^t e^{2z_1 t'} \delta(t' - \omega/\beta) dt' \quad (22)$$

$$= \frac{2}{|\beta|} u(\tau) e^{-2z_1(\tau)} \quad (23)$$

where  $\tau$  is defined as in (6).

The solution to the other equations is obtained with the same technique. Here  $W_2$  and  $W_3$  are shown, while  $W_{x,x}$  is reported in (5)

$$W_2 = \frac{2}{|\beta|} \frac{1}{z_2 - z_1} u(\tau) \left[ e^{-2z_1(\tau)} - e^{-2z_2(\tau)} \right] \quad (24)$$

$$W_3 = \frac{2}{|\beta|} \frac{1}{z_2 - z_1} u(\tau) \quad (25)$$

$$\left[ \frac{1}{\bar{z}_1 - z_1} \left( e^{-2z_1(\tau)} - e^{-2\bar{z}_1(\tau)} \right) - \frac{1}{\bar{z}_1 - z_2} \left( e^{-2z_2(\tau)} - e^{-2\bar{z}_1(\tau)} \right) \right] \quad (26)$$

When solving those equations, the evaluation of the following integral is always encountered

$$\int_{-\infty}^t u(t' - \omega/\beta) e^{2(z_2 - z_1)t'} dt' = \quad (27)$$

$$u(t - \omega/\beta) \int_{\omega/\beta}^t e^{2(z_2 - z_1)t'} dt' = \quad (28)$$

$$u(t - \omega/\beta) \frac{e^{2(z_2 - z_1)t} - e^{2(z_2 - z_1)\omega/\beta}}{2(z_2 - z_1)} \quad (29)$$

## REFERENCES

- [1] N. F. Barber and F. Ursell, "The response of a resonant system to a gliding tone," *Phil. Mag.*, vol. 39, pp. 345-361, 1948.
- [2] L. Cohen, *Time-Frequency Analysis*, Prentice-Hall, 1995.
- [3] L. Galleani, L. Cohen, "Dynamics Using the Wigner Distribution", *to appear in ICPR 2000*, September 3-8 2000, Barcelona - SPAIN
- [4] G. Hok, "Response of linear Resonant systems to Excitation of a frequency Varying Linearly with Time" *Journal of Applied Physics*, vol. 19, pp. 242-250, 1948.
- [5] W. T. Thomson, *Theory of Vibrations with Applications*, Chapman and Hall, 1983.
- [6] T.A.C.M. Claasen and W.F.G. Mecklenbrauker, "The Wigner Distribution etc..." *Philips Journal of Research*, vol. 35, No. 3, 1980.

# BASELINE AND DISTRIBUTION ESTIMATES OF COMPLICATED SPECTRA

David J. Thomson

Bell Labs, Murray Hill, New Jersey 07974

## ABSTRACT

In processes with many lines, such as those encountered in climate, space physics, and communications systems, it is often found that the underlying base spectrum<sup>1</sup> also has a complicated shape. In this paper I discuss some methods to estimate the base, and detail spectra by using the probability distributions of such spectra. The methods include a robust estimate of the central part of a mixture of central and noncentral chi-square distributions and, second, a direct estimate of the mixing and noncentrality parameters made by minimizing the Kolmogorov-Smirnov  $D$  statistic between the empirical and theoretical cumulative distribution functions.

## 1. INTRODUCTION

The problem considered in this paper is that of analyzing the incoherent spectrum of data when there are many, possibly thousands, of lines within the Nyquist band. Typically these are not strict sinusoids, but have unknown narrow-band modulations. The problems of interest are separating the "base" and "detail" spectra, estimating the number of lines, and the power in these lines. The base spectrum is unknown and assumed to have a moderately complicated, but smooth, shape.

In the problems considered here, the basic data is assumed to consist of spectrum estimates over a range of frequencies that, apart from an overall "red" shape, have a mixture of central and non-central  $\chi^2_\nu$  distributions.  $\nu$ , the degrees-of-freedom, is known and the scale factors for the two components of the mixture are assumed to be the same. The mixing fraction, scale, and noncentrality parameters are to be estimated. Two estimation procedures are described here. The first is a simple robust estimate made by taking a scaled quantile of the estimates in a given frequency range. The second procedure uses a goodness-of-fit test and

<sup>1</sup>I have termed the two parts of the spectrum *base* and *detail* as opposed to, for example, fit and residual. This is to emphasize that the two components are both of interest, the base varying more slowly in frequency than the detail. In addition, "residual" is often dismissed as "noise" and usually implies subtraction, not division.

chooses the parameters that minimize the misfit between the observed and theoretical cumulative distribution functions. The Kolmogorov-Smirnov  $D$  statistic is used as a measure of misfit in these examples.

An example of such a problem is that of solar  $g$ -modes in the interplanetary magnetic field (IMF). A few years ago in Thomson *et al.* (1995), we proposed that the observed fluctuations of energetic particles in the solar wind were a result of solar  $g$ -(gravity) and  $p$ -(pressure) modes thus contradicting the prevailing opinion that the fluctuations were from turbulence. In theory,  $g$ -modes have extremely high  $Q$ 's,  $\sim 10^{11}$ , Bahcall and Kumar (1993), but whether the frequencies are stable, or modulated by the solar cycle, is unknown. In either case the high  $Q$ 's imply that the modes lose very little energy either to dissipation or radiation so that detection is not simple. Because magnetic fields are fundamental in space physics, establishing the presence and characteristics of modes in the IMF is crucial for proper understanding. The data used here are one-hour averages of the normal component of the interplanetary magnetic field measured by the Ulysses spacecraft, Balogh *et al.* (1992). These were measured while Ulysses was near the ecliptic plane between day 298 of 1990, just after launch, until day 33 of 1992, just before the spacecraft's Jupiter encounter. The 11,157 hourly measurements span a radial distance of 1 to 5.3 Astronomical Units. The data were multiplied by heliocentric distance to remove the radial dependency.

## 2. DATA ASSUMPTIONS

I assume that the data contains *narrow band* quasi-deterministic components, typically sinusoidal components that have slow amplitude or phase modulations.<sup>2</sup> Denote such a modulation shape by  $M_m(t)$ , standard-

<sup>2</sup>In addition to the unknown solar processes involved in transferring small mechanical motions in the core of the sun through the convection zone, photosphere, and corona into the interplanetary magnetic field, the spacecraft's motion along its orbit causes additional confounding effects. Because the modal amplitudes are spherical harmonics, the observed amplitude will depend on both the spacecraft's heliographic latitude and radius and its changing velocity will give a Doppler shift.

ized to have unit energy,

$$\sum_{t=0}^{N-1} |M_m(t)|^2 = 1 \quad (1)$$

and assume that the process is of the form

$$x(t) = \zeta(t) + \sum_{m,n} \mu_{m,n} e^{i2\pi f_{m,n} t} M_m(t) \quad (2)$$

where  $\zeta(t)$  is a nondeterministic process with a relatively slowly-varying spectrum. I assume that  $\zeta(t)$  is independent of the line components. There are an unknown number of modulation signals and a large number of signal frequencies  $f_{m,n}$ .

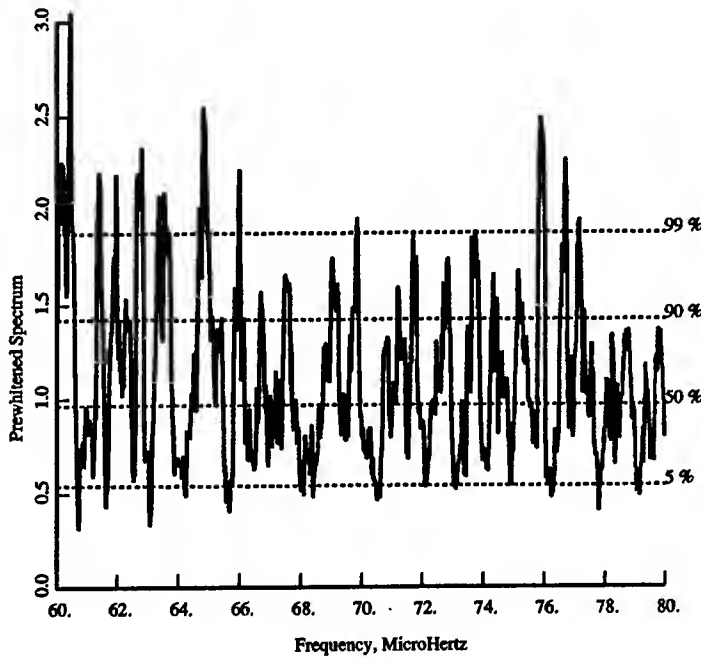


Figure 1: A portion of the estimated detail spectrum,  $\hat{S}(f)$ , of the Ulysses normal magnetic field. The spectrum has been levelled, by the estimate shown in Figure 4, to have a unit base power level. There are 20 peaks above the 99% level, cumulatively 7.9% of the estimates compared to 4.9% below the 5% level.

### 3. MULTITAPER ESTIMATES

In the usual formulation of multitaper spectrum estimation, Thomson (1982), one is given a sample of  $N$  equally-spaced data  $x(t)$ . One chooses a bandwidth  $W$  from exploratory data analysis and, in this case, from the theoretical spacings of  $g$ -modes, Guenther *et al.* (1992). One then computes the  $K = 2NW$  Slepian

sequences,  $v_t^k(N, W)$ , or discrete prolate spheroidal sequences, Slepian (1978), and the windowed Fourier transforms or eigencoefficients,

$$y_k(f) = \sum_{t=0}^{N-1} e^{-i2\pi ft} v_t^{(k)}(N, W) x(t). \quad (3)$$

I used a time-bandwidth product  $NW = 6$  and  $K = 10$  windows. Because the higher-order coefficients, those for  $k \lesssim K - 1$ , are more susceptible to broad band bias than those for small  $k$ , one forms an estimate,  $\hat{x}_k(f)$ , of the ideal eigencoefficients that would be obtained if the frequency band  $(f - W, f + W)$  could be observed in isolation. (See Thomson §V of 1982 or Thomson §3.3 of 2000 for details.) The canonical multitaper spectrum estimate is then

$$\hat{S}_x(f) = \frac{1}{K} \sum_{k=0}^{K-1} |\hat{x}_k(f)|^2. \quad (4)$$

The eigencoefficients of the unknown modulation signals  $M_m(t)$ ,

$$M_{m,k}(f) = \sum_{t=0}^{N-1} v_t^{(k)} M_m(t) e^{i2\pi ft}, \quad (5)$$

become, in the frequency domain, a convolution of the Fourier transforms of  $M_m(t)$  with the Slepian functions so the apparent bandwidth of a line will increase beyond the  $2W$  of the Slepian functions by the bandwidth of the modulating signal. Thus the assumption that the  $M_m$ 's are narrow-band can be checked by testing that the observed linewidths are close to  $\pm W$ . A histogram of the peak widths was made, Figure 2. Peaks exceeding the 95% significance level of the  $\chi_{20}^2$  distribution in the detail spectrum, Figure 1, were used, with the widths measured at the 90% point. This histogram has a sharp peak located close to the width expected with pure sinusoids, so one can conclude that almost all the energy of a line at frequency  $f$  is contained in the frequency band  $(f - W, f + W)$ .

Now consider a multitaper spectrum estimated at one of these frequencies, say  $f_0 = f_{m,n}$ . The eigencoefficients are

$$y_k(f_0) = \zeta_k(f_0) + \mu_{m,n} M_{m,k} \quad (6)$$

and the simple spectrum estimate  $\hat{S}_x(f_0)$  will have a *non-central* chi-square distribution with  $2K$  degrees-of-freedom. If one denotes the spectrum of the base process by  $S_\zeta(f)$  and defines the detail spectrum by

$$S(f) = \frac{S_x(f)}{S_\zeta(f)}$$

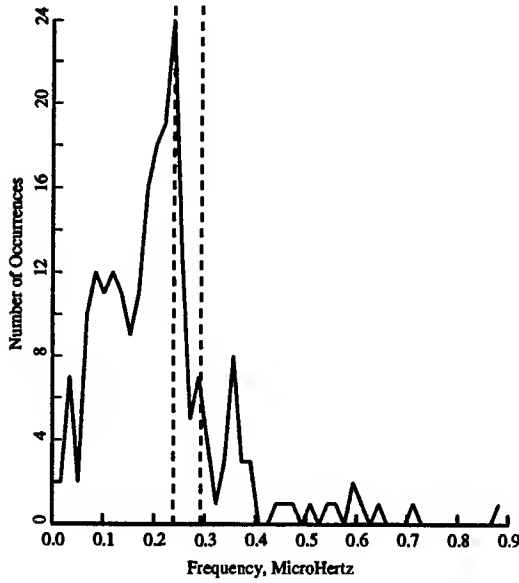


Figure 2: Histogram of peak widths from the full 0 to 140  $\mu\text{Hz}$  frequency range. The dotted vertical lines are the base width,  $2W \approx 0.29\mu\text{Hz}$  of the spectral window and,  $0.23\mu\text{Hz}$ , its approximate width where the measurements are made.

then, between lines,  $2K\hat{S}(f)$  will have a *central* chi-square distribution with  $\nu = 2K$  degrees-of-freedom. Similarly, at line frequencies, the distribution will be *non-central*  $\chi^2_{2K}$ . Using (1) and the narrow-band properties of the modulation, the non-centrality parameter is

$$\lambda = \frac{2K|\mu_{m,n}|^2}{S_c(f_0)}. \quad (7)$$

Thus, depending on whether a frequency  $f$  happens to fall on a line or between lines, the estimated spectrum  $\hat{S}_x(f)$  will have either a central or noncentral  $\chi^2_\nu$  distribution and, overall, a mixture.

There are, consequently, two distinct problems: *first* estimating the baseline spectrum,  $S_c(f_0)$  and; *second* estimating the non-centrality parameter,  $\lambda$ .

It should be emphasized that the base spectrum *must* be removed before distribution tests are attempted because, if not, the usual red spectral shape simply causes the distribution tests to be unreliable. Because noncentral chi-square distributions may be approximated with mixtures of central chi-squares, Johnson *et al.* (1995, Ch. 29), the mixture resulting from an incorrectly estimated base spectrum may thus be mistaken for a noncentral distribution. Further, if the base spectrum is allowed to vary too rapidly, moderate peaks can be suppressed so the noncentral component will not be detected. Consequently, both false detection and rejection failures are possible.

It can be asked why maximum-likelihood estimates are not used here, and there are several answers. First, MLE's for noncentral chi-square distributions are at both more complicated and can have poorer performance than moment estimates, Alam and Saxena (1982). Second, MLE's of chi-square *mixtures* do not appear to have been studied. Third, the computation burden of the MLE's is at least as severe as that of the direct goodness-of-fit tests. Fourth, the robust estimate works when there are several noncentral distributions in the mixture, whereas, with a MLE, the number of different distributions would have to be known or estimated.

#### 4. A SIMPLE ROBUST ESTIMATE

Suppose we have a set of  $J$  samples of a mixture distribution,  $s_j$ ,  $j = 1, \dots, J$ . The distribution of the individual  $s_j$ 's is, with probability  $(1 - \epsilon)$ , central  $\chi^2_\nu$  or, with probability  $\epsilon$ , non-central  $\chi^2_\nu(\lambda)$ . All the  $s_j$ 's have a common unknown scale,  $a$ , so the central component is the same in either case.  $\nu$ , is known so the probability density function for  $a = 1$  is

$$p_\epsilon(s|\nu, \lambda, \epsilon) = (1 - \epsilon)p_c(s|\nu) + \epsilon p_{nc}(s|\nu, \lambda) \quad (8)$$

where  $p_c$  and  $p_{nc}$  denote the standard  $\chi^2_\nu$  central and non-central densities respectively, Lancaster (1969). The corresponding cumulative distributions,  $P$ , and quantiles,  $Q$ , are similarly denoted,

$$\int_0^{Q_\epsilon(p)} p_c(s|\nu) ds = P_c(Q_\epsilon(p)|\nu) = p.$$

The expected value of the mixture is

$$E s_j = a\{(1 - \epsilon)\nu + \epsilon(\nu + \lambda)\} = a\{\nu + \epsilon\lambda\} \quad (9)$$

and the problem is to estimate  $a$ ,  $\epsilon$ , and  $\lambda$ .

Robust estimates of  $\chi^2$  distributions do not appear to be a well-studied problem. For example, neither "robust" nor "non-central" is indexed in Bowman and Shenton (1988). Consider a simple estimate based on order statistics: Denote the sorted observations by  $s_{(j)}$  with

$$s_{(1)} \leq s_{(2)} \leq \dots \leq s_{(J)}$$

and consider an estimate based on the  $p^{\text{th}}$  quantile,  $Q_\epsilon(p, \nu)$  of the *central* distribution. From the empirical cumulative probability distribution of the  $j^{\text{th}}$  sample point,  $p_j = (2j - 1)/(2J)$ , define a scale factor  $\beta_j = 1/Q_\epsilon(p_j, \nu)$  so the estimate is

$$\hat{a}(j) = \beta_j s_{(j)}. \quad (10)$$

For given  $\epsilon$  and  $\lambda$  and taking  $Q_\epsilon(j)$  as the  $p^{\text{th}}$  quantile of the mixture distribution, the variance, using methods from Kendall and Stuart (1963), is

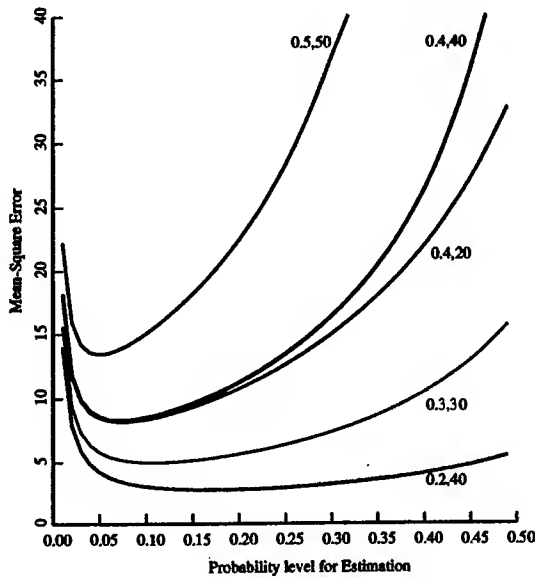


Figure 3: The mean-squared-error for a  $\chi^2_{20}$  mixture with various values of  $\epsilon$  and  $\lambda$  vs the probability level. The minimum MSE for cases of interest typically occurs for  $p$  in the 5 to 10 percent range.

$$\text{Var}\{\hat{a}(j)\} \approx \frac{p_j(1-p_j)\nu^2}{Jp_c^2(Q_c(j))Q_c(j)^2}$$

and, similarly, the bias is  $\beta_j Q_c(j) - \nu$ . Evaluating the mean-square-error, Figure 3, one finds that, for typical values of  $\epsilon$  and  $\lambda$ , the optimum quantile is rather low, usually between the 5 and 10% points. By most standards this is surprising, but, on considering that much less work has been done on robust estimates in nonsymmetric distributions than on the symmetric case, less so than at first glance. Earlier work, Thomson (1977, Part II, §V), showed similar robust estimates for standard  $\chi^2_2$  estimates that also trimmed approximately the top third of the population, but with the primary motivation of eliminating spectra estimated from contaminated data, not for estimating lines. They were, however, see *e.g.* Figures 7, 8, and 11 of Thomson (1977), also effective for finding lines that simple section averages missed.

## 5. ESTIMATING THE NOISE SPECTRUM

The base spectrum  $S_c(f)$  was estimated by, first, dividing the raw spectrum estimate,  $\hat{S}_x(f)$ , with an autoregressive fit,  $S_{ar}(f)$ , to obtain an approximately white intermediate spectrum  $\hat{S}_i(f) = \hat{S}_x(f)/S_{ar}(f)$ . Next, the robust estimate (10) was slid along  $\hat{S}_i(f)$  to get an estimate of the central component,  $\hat{S}_c(f)$ . Finally, one takes  $S_{ar}(f) \times \hat{S}_c(f)$  as a robust estimate

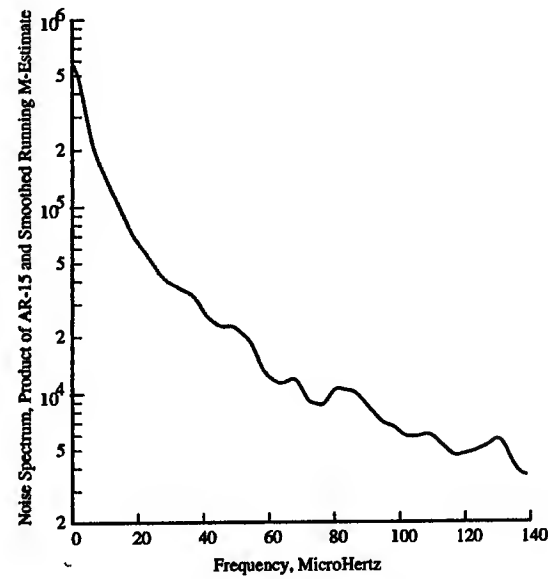


Figure 4: The base spectrum,  $\hat{S}_c(f)$ , estimated by taking an initial AR-15 spectrum times a smoothed running quantile estimate. The details are statistically significant.

of  $\hat{S}_c(f)$ . Figure 1 is a plot of the detail spectrum,  $\hat{S}(f) = \hat{S}_x(f)/\hat{S}_c(f)$ . One implication of the lowest mean-square-error occurring near the 5% point of the distribution implies that moderately large samples are required to apply this method<sup>3</sup>. Exploratory data analysis with various estimation spans,  $J$ , or, in bandwidth,  $(J-1)\Delta f$ , showed that, for this data  $J \propto 2.3\mu\text{Hz}$  was a reasonable size. This was determined by requiring that, on average, the variation in  $\hat{S}(f)$ , as measured by the ratio of the 90% to 10% points, does not drop below that expected for a central  $\chi^2_2$ . The range between the 10% and 90% points does not depend strongly on sample size, so testing for overfitting the base spectrum was relatively easy. Figure 4 shows an estimate,  $\hat{S}_c(f)$ . The original spectrum estimate,  $\hat{S}_x(f)$ , was divided by this base estimate to get the detail spectrum part of which is shown in Figure 1. The ripples in the base spectrum are large enough to rule out the simple power-law spectrum predicted by turbulence theory.

## 6. FITS TO DISTRIBUTION

Fitting a mixture density is a nonlinear procedure. Because probability density functions must be positive, even partial linearization for  $\epsilon$  can give unacceptable results. The procedure adopted was to use Brent, 's 1973 algorithm, FMIN, to minimize the misfit between

<sup>3</sup>Estimates at several probability levels, using (10), were made and compared, and a smoothed average of those at the 5 and 10% levels was used.

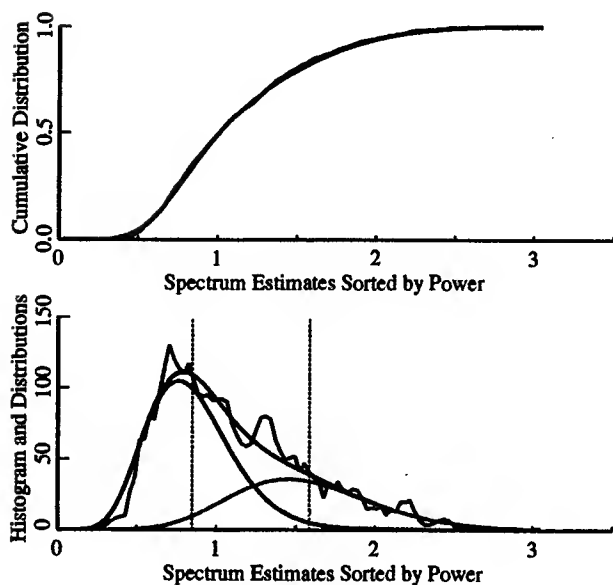


Figure 5: Empirical and the best fitting mixture cumulative distributions for the detail spectrum shown in Figure 1. The maximum discrepancy is  $D = 0.0191$ , near the center of the distribution. The lower panel shows a histogram of the data, the central and noncentral probability densities, and their sum, all scaled by sample size. The estimated mean of the central distribution is at a level of 0.8, showing that the base estimate was about 20% too high.

the empirical cdf,  $(j - \frac{1}{2})/J$  and the theoretical cdf,  $P_\epsilon(s_j|\nu, \epsilon, \lambda)$ . For trial values of  $\epsilon$  and  $\lambda$  the scale,  $a$ , was estimated by equating the theoretical mean (9) to the observed average<sup>4</sup>. A goodness-of-fit test, the Kolmogorov-Smirnov  $D$  statistic was used as a measure of misfit. Figure 5 shows this procedure applied to the detail spectrum of Figure 1. The fit is extremely good and, in this example, the estimated non-centrality parameter is  $\lambda \sim 0.8\nu \approx 0.36$  and about 30 percent of the total power is in the non-central component.

## References

- Alam, K. and Saxena, L. (1982). Estimation of the noncentrality parameter of a chi-square distribution. *Annals of Statistics*, **10**, 1012–1016.
- Bahcall, J. N. and Kumar, P. (1993).  $g$ -modes and the solar neutrino problem. *Astrophys. J.*, **409**, L73–L76.
- Balogh, A., Forsyth, R. J., Hedgecock, P. C., Marquedant, R. J., Smith, E. J., Southworth, D. J., and Tsurutani, B. T. (1992). The magnetic field investigation on the

Ulysses mission: Instrumentation and preliminary scientific results. *Astron. and Astrophysics, Suppl. Series*, **92**, 221–236.

- Bennett, B. M. (1955). Note on the moments of the logarithmic noncentral  $\chi^2$  and  $z$  distributions. *Annals of the Institute of Statistical Mathematics*, **7**, 57–61.
- Bowman, K. O. and Shenton, L. R. (1988). *Properties of Estimators for the Gamma Distribution*. Marcel Dekker, New York.
- Brent, R. P. (1973). *Algorithms for Minimization without Derivatives*. Prentice-Hall, Englewood Cliffs, NJ.
- Guenther, D. B., Demarque, P., Pinsonneault, M. H., and Kim, Y.-C. (1992). Standard solar models. ii.  $g$ -modes. *Astrophys. J.*, **392**, 328–336.
- Johnson, N. L., Kotz, S., and Balakrishnan, N. (1995). *Continuous Univariate Distributions*. John Wiley and Sons, New York, second edition.
- Kendall, M. G. and Stuart, A. (1963). *The Advanced Theory of Statistics*. Hafner, New York.
- Lancaster, H. O. (1969). *The Chi-squared Distribution*. John Wiley and Sons, New York.
- Slepian, D. (1978). Prolate spheroidal wave functions, Fourier analysis, and uncertainty V: the discrete case. *Bell System Tech. J.*, **57**, 1371–1429.
- Thomson, D. J. (1977). Spectrum estimation techniques for characterization and development of WT4 waveguide. *Bell System Tech. J.*, **56**, Part I, 1769–1815, Part II, 1983–2005.
- Thomson, D. J. (1982). Spectrum estimation and harmonic analysis. *Proc. IEEE*, **70**, 1055–1096.
- Thomson, D. J. (2000). Multitaper analysis of nonstationary and nonlinear time series data. In W. Fitzgerald, R. Smith, A. Walden, and P. Young, editors, *Nonlinear and Nonstationary Signal Processing*. Cambridge Univ. Press.
- Thomson, D. J., MacLennan, C. G., and Lanzerotti, L. J. (1995). Propagation of solar oscillations through the interplanetary medium. *Nature*, **376**, 139–144.

<sup>4</sup>There is some indication that the log moments may also be good estimators for this. See Bennett (1955) for expected values.

# DISTRIBUTED SOURCE LOCALIZATION WITH MULTIPLE SENSOR ARRAYS AND FREQUENCY-SELECTIVE SPATIAL COHERENCE

Richard J. Kozick

Bucknell University  
Lewisburg, PA 17837  
kozick@bucknell.edu

Brian M. Sadler

Army Research Laboratory  
Adelphi, MD 20783  
bsadler@arl.mil

## ABSTRACT

Multiple sensor arrays distributed over a planar region provide the means for highly accurate localization of the  $(x, y)$  position of a source. In some applications, such as microphone arrays receiving aeroacoustic signals from ground vehicles, random fluctuations in the air lead to frequency-selective coherence of the signals that arrive at widely separated arrays. We present a performance analysis for localization of a wideband source using multiple sensor arrays. The wavefronts are modeled with perfect spatial coherence over individual arrays and with frequency-selective coherence between distinct arrays. The sensor signals are modeled as wideband Gaussian random processes, and we study the Cramer-Rao bound (CRB) on source localization accuracy for varying levels of signal coherence and for processing schemes with different levels of complexity. We show that significant improvements in source localization accuracy are possible when partial signal coherence from array to array is exploited. Further, we show that a distributed processing scheme involving bearing estimation at the individual arrays and time-delay estimation between pairs of sensors performs nearly as well as the optimum scheme that jointly processes the signals from all sensors. Results based on measured aeroacoustic data are included to illustrate frequency-selective signal coherence at distributed arrays.

## 1. INTRODUCTION

We are concerned with estimating the location  $(x_s, y_s)$  of a wideband source using multiple sensor arrays that are distributed over an area. We consider schemes that distribute the processing between the individual arrays and a fusion center in order to limit the communication bandwidth between arrays and fusion center. Triangulation is a standard approach for source localization with multiple sensor arrays. Each array estimates a bearing and transmits the bearing to the fusion center, which combines the bearings to estimate the source location  $(x_s, y_s)$ . Triangulation is characterized by low communication bandwidth and low complexity, but it ignores *coherence* that may be present in the wavefronts that are received at distributed arrays. In this paper, we investigate new approaches for source localization with multiple arrays that exploit partial coherence of the wavefronts at distributed arrays. We show that the Cramer-Rao lower bound (CRB) on estimating the source location is significantly reduced when coherence from array to array is ex-

ploited. We also evaluate the performance of suboptimum source localization methods that employ distributed processing to reduce the communication bandwidth between the arrays and the fusion center. Results are presented from processing measured aeroacoustic data to illustrate signal coherence at distributed arrays.

Previous work on source localization with aeroacoustic arrays has focused on angle of arrival estimation with a *single* array [1]. The problem of imperfect spatial coherence in the context of narrowband angle-of-arrival estimation with a single array has been studied in [2]-[5]. Pauraj and Kailath [2] presented a MUSIC algorithm that incorporates the nonideal spatial coherence, assuming that the coherence variation is known. Gershman et al. [3] provided a procedure to jointly estimate the spatial coherence loss and the angles of arrival. Song and Ritcey [4] provide maximum-likelihood (ML) methods for estimating the parameters of a coherence model and the angles of arrival, and Wilson [5] incorporates physical models for the spatial coherence. The problem of decentralized array processing has been studied in [6]-[8]. Wax and Kailath [6] present subspace algorithms for narrowband signals and distributed arrays, assuming perfect spatial coherence across each array but neglecting the spatial coherence between arrays. Weinstein [7] presents performance analysis for pairwise processing the wideband sensor signals from a single array and shows negligible loss in localization accuracy when the SNR is high. Stoica, Nehorai, and Soderstrom [8] consider ML angle of arrival estimation with a large, perfectly coherent array that is partitioned into subarrays.

## 2. DATA MODEL

A model is formulated in this section for the signals received by the sensors in distributed arrays. Consider a single source that is located at coordinates  $(x_s, y_s)$  in the  $(x, y)$  plane. Then  $H$  arrays are distributed in the same plane, as illustrated in Figure 1. The signals measured at the distributed sensor arrays are modeled as jointly Gaussian wideband random processes. The model is very general, and it accounts for propagation effects between the source and the distributed arrays, including frequency-selective spatial coherence and different signal power levels received at each array. The spatial coherence of the wavefronts is modeled as being perfect over each individual array but variable between distinct arrays. This idealization allows us to study

the effect of varying coherence between arrays on source localization accuracy. Physical modeling of frequency selective coherence is discussed in [9]. The power spectral density of the source is arbitrary, allowing a range of cases to be modeled including narrowband sources and sums of harmonics, as well as wideband sources with continuous power spectra.

Each array  $h \in \{1, \dots, H\}$  contains  $N_h$  sensors, and has a reference sensor located at coordinates  $(x_h, y_h)$ . The location of sensor  $n \in \{1, \dots, N_h\}$  is at  $(x_h + \Delta x_{hn}, y_h + \Delta y_{hn})$ , where  $(\Delta x_{hn}, \Delta y_{hn})$  is the relative location with respect to the reference sensor. If  $c$  is the speed of propagation, then the propagation time from the source to the reference sensor on array  $h$  is

$$\tau_h = \frac{r_h}{c} = \frac{1}{c} [(x_s - x_h)^2 + (y_s - y_h)^2]^{1/2}. \quad (1)$$

We will assume that the wavefronts are well approximated by plane waves over the aperture of individual arrays. The propagation time from the source to sensor  $n$  on array  $h$  will be expressed by  $\tau_h + \tau_{hn}$ , where

$$\begin{aligned} \tau_{hn} &\approx -\frac{1}{c} \left[ \frac{x_s - x_h}{r_h} \Delta x_{hn} + \frac{y_s - y_h}{r_h} \Delta y_{hn} \right] \\ &= -\frac{1}{c} [(\cos \phi_h) \Delta x_{hn} + (\sin \phi_h) \Delta y_{hn}], \end{aligned} \quad (2)$$

where  $\tau_{hn}$  is the propagation time from the reference sensor on array  $h$  to sensor  $n$  on array  $h$ , and  $\phi_h$  is the bearing of the source with respect to array  $h$ . Note that while the far-field approximation (2) is reasonable over individual array apertures, the wavefront curvature that is inherent in (1) must be retained in order to accurately model the (possibly) wide separation between arrays.

The time signal received at sensor  $n$  on array  $h$  due to the source will be represented as  $s_h(t - \tau_h - \tau_{hn})$ , where the vector of signals  $\mathbf{s}(t) = [s_1(t), \dots, s_H(t)]^T$  received at the  $H$  arrays are modeled as real-valued, continuous-time, zero-mean, wide-sense stationary, Gaussian random processes with  $-\infty < t < \infty$ . These processes are fully specified by the  $H \times H$  cross-correlation function matrix

$$\mathbf{R}_s(\tau) = E\{\mathbf{s}(t + \tau) \mathbf{s}(t)^T\}, \quad (3)$$

where  $E$  denotes expectation, superscript  $T$  denotes transpose, and we will later use the notation superscript  $*$  and superscript  $H$  to denote complex conjugate and conjugate transpose, respectively. The  $(g, h)$  element in (3) is the cross-correlation function

$$r_{s,gh}(\tau) = E\{s_g(t + \tau) s_h(t)\} \quad (4)$$

between the signals received at arrays  $g$  and  $h$ . The correlation functions (3) and (4) are equivalently characterized by their Fourier transforms, which are the cross-spectral density functions and matrix

$$\begin{aligned} G_{s,gh}(\omega) &= \mathcal{F}\{r_{s,gh}(\tau)\} = \int_{-\infty}^{\infty} r_{s,gh}(\tau) \exp(-j\omega\tau) d\tau \\ \mathbf{G}_s(\omega) &= \mathcal{F}\{\mathbf{R}_s(\tau)\}. \end{aligned} \quad (5)$$

The diagonal elements  $G_{s,hh}(\omega)$  of (5) are the power spectral density (PSD) functions of the signals  $s_h(t)$ , and hence

they describe the distribution of average signal power with frequency. The model allows the average signal power to vary from one array to another. Indeed, the PSD may even vary from one array to another to reflect propagation differences, source aspect angle differences, and other effects that lead to coherence degradation in the signals at distributed arrays.

Let us elaborate the definition and the meaning of *coherence* between the signals  $s_g(t)$  and  $s_h(t)$  received at distinct arrays  $g$  and  $h$ . In general, the cross-spectral density function (5) can be expressed in the form

$$G_{s,gh}(\omega) = \gamma_{s,gh}(\omega) [G_{s,gg}(\omega) G_{s,hh}(\omega)]^{1/2}, \quad (6)$$

where  $\gamma_{s,gh}(\omega)$  is the spectral coherence function, which has the property  $0 \leq |\gamma_{s,gh}(\omega)| \leq 1$ . The coherence function  $\gamma_{s,gh}(\omega)$  is generally complex-valued, but we will model it as real-valued. This is a reasonable assumption for acoustic propagation environments in which the loss of coherence is due to random changes in the apparent source location, as long as the change in apparent source location is the same at both arrays  $g$  and  $h$  [5, 9].

We model the signal received at sensor  $n$  on array  $h$  as a sum of the delayed source signal and noise,

$$z_{hn}(t) = s_h(t - \tau_h - \tau_{hn}) + w_{hn}(t), \quad (7)$$

where the noise signals  $w_{hn}(t)$  are modeled as real-valued, continuous-time, zero-mean, wide-sense stationary, Gaussian random processes that are uncorrelated at distinct sensors. The noise correlation properties are

$$E\{w_{gm}(t + \tau) w_{hn}(t)\} = r_w(\tau) \delta_{gh} \delta_{mn}, \quad (8)$$

where  $r_w(\tau)$  is the noise autocorrelation function, and the noise power spectral density is  $G_w(\omega) = \mathcal{F}\{r_w(\tau)\}$ . We then collect the observations at each array  $h$  into  $N_h \times 1$  vectors  $\mathbf{z}_h(t) = [z_{h1}(t), \dots, z_{hN_h}(t)]^T$  for  $h = 1, \dots, H$ , and we further collect the observations from the  $H$  arrays into a  $(N_1 + \dots + N_H) \times 1$  vector

$$\mathbf{Z}(t) = [\mathbf{z}_1(t)^T \dots \mathbf{z}_H(t)^T]^T. \quad (9)$$

The elements of  $\mathbf{Z}(t)$  in (9) are zero-mean, wide-sense stationary, Gaussian random processes. We can express the cross-spectral density matrix of  $\mathbf{Z}(t)$  in a convenient form with the following definitions. The array manifold for array  $h$  at frequency  $\omega$  is

$$\begin{aligned} \mathbf{a}_h(\omega) &= \begin{bmatrix} \exp(-j\omega\tau_{h1}) \\ \vdots \\ \exp(-j\omega\tau_{hN_h}) \end{bmatrix} \\ &= \begin{bmatrix} \exp[j\frac{\omega}{c} ((\cos \phi_h) \Delta x_{h1} + (\sin \phi_h) \Delta y_{h1})] \\ \vdots \\ \exp[j\frac{\omega}{c} ((\cos \phi_h) \Delta x_{hN_h} + (\sin \phi_h) \Delta y_{hN_h})] \end{bmatrix}, \end{aligned} \quad (10)$$

using  $\tau_{hn}$  from (2) and assuming that the sensors have omnidirectional response to sources in the plane of interest. Let us define the relative time delay of the signal at arrays  $g$  and  $h$  as  $D_{gh} = \tau_g - \tau_h$ , where  $\tau_h$  is defined in (1). Then the cross-spectral density matrix of  $\mathbf{Z}(t)$  in (9) has the form



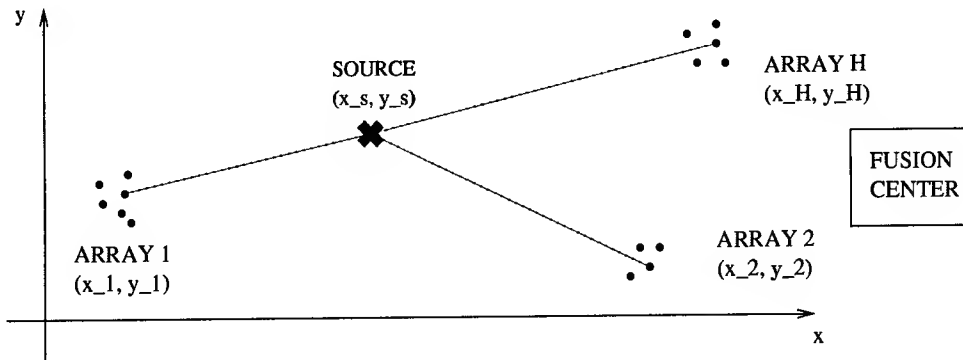


Figure 1: Geometry of source location and  $H$  distributed sensor arrays. A communication link is available between each array and the fusion center.

shown in (11) in Figure 2. Recall that the source cross-spectral density functions  $G_{s,gh}(\omega)$  in (11) can be expressed in terms of the spectral coherence  $\gamma_{s,gh}(\omega)$  using (6).

Note that (11) depends on the source location parameters  $(x_s, y_s)$  through  $\mathbf{a}_h(\omega)$  and  $D_{gh}$ . However, (11) points out that the observations are also characterized by the bearings  $\phi_1, \dots, \phi_H$  to the source from the individual arrays and the relative time delays  $D_{gh}$  between pairs of arrays. Therefore, one way to estimate the source location  $(x_s, y_s)$  is to first estimate the bearings  $\phi_1, \dots, \phi_H$  and the pairwise time delays  $D_{gh}$ .

### 3. CRBS ON LOCALIZATION ACCURACY

The problem of interest is to estimate the source location parameter vector  $\Theta = [x_s, y_s]^T$  using  $T$  samples of the sensor signals  $\mathbf{Z}(0), \mathbf{Z}(T_s), \dots, \mathbf{Z}((T-1) \cdot T_s)$ , where  $T_s$  is the sampling period. Let us denote the sampling rate by  $f_s = 1/T_s$  and  $\omega_s = 2\pi f_s$ . We will assume that the continuous-time random processes  $\mathbf{Z}(t)$  are band-limited, and that the sampling rate  $f_s$  is greater than twice the bandwidth of the processes. Then Friedlander [10] has shown that the Fisher information matrix (FIM)  $\mathbf{J}$  for the parameters  $\Theta$  based on the samples  $\mathbf{Z}(0), \mathbf{Z}(T_s), \dots, \mathbf{Z}((T-1) \cdot T_s)$  has elements  $J_{ij}$  shown in (12) in Figure 2. The CRB matrix  $\mathbf{C} = \mathbf{J}^{-1}$  then has the property that the covariance matrix of any unbiased estimator  $\hat{\Theta}$  satisfies  $\text{Cov}(\hat{\Theta}) - \mathbf{C} \geq 0$ , where  $\geq 0$  means that  $\text{Cov}(\hat{\Theta}) - \mathbf{C}$  is positive semidefinite. The CRB provides a lower bound on the performance of any unbiased estimator. Equation (12) provides a convenient way to compute the FIM for the distributed sensor array model. It provides a powerful tool for evaluating the impact that various parameters have on source localization accuracy. Parameters of interest include the spectral coherence between distributed arrays, the signal bandwidth and power spectrum, the array placement geometry, and the SNR. The FIM in (12) is not easily evaluated analytically, but it is readily evaluated numerically for cases of interest. The FIM expression (12) can be specialized for two important cases. With  $H = 2$  arrays containing  $N_1 = N_2 = 1$  sensor each, we obtain a generalization of the classic time delay estimation problem [11] with *partial* signal coherence at the sensors. For arbitrary number of arrays  $H$  and  $N_1, \dots, N_H$ ,

we can specialize (12) for sources with a *narrowband* power spectrum.

The CRB based on (12) provides a performance bound on source location estimation methods that jointly process all the data from all the sensors. Such processing provides the best attainable results, but it also requires significant communication bandwidth to transmit data from the individual arrays to the fusion center. We have developed performance bounds for schemes that perform bearing estimation at the individual arrays in order to reduce the required communication bandwidth to the fusion center. These CRBs facilitate a study of the tradeoff between source location accuracy and communication bandwidth between the arrays and the fusion center. Two methods are considered [12]:

1. Ordinary triangulation, where each array estimates the source bearing and transmits the bearing estimate to the fusion center. This approach does not exploit wavefront coherence between the distributed arrays, but it minimizes the communication bandwidth between the arrays and the fusion center.
2. Each array estimates the source bearing and transmits the bearing estimate to the fusion center. In addition, the raw data from *one sensor* in each array is transmitted to the fusion center. The fusion center then estimates the propagation time delay between pairs of distributed arrays, and triangulates these time delay estimates with the bearing estimates to localize the source.

Method 2 performs nearly as well as optimum joint processing if the SNR is high enough.

### 4. EXAMPLES

We present an example that illustrates the potential improvement in source localization accuracy when coherence between the distributed arrays is exploited. Consider a scenario with  $H = 3$  arrays, where the individual arrays are identical and contain  $N_1 = N_2 = N_3 = 7$  sensors. Each array is circular and has 4-ft radius, with six sensors equally spaced around the perimeter and one sensor in the center. Narrowband processing in a 1-Hz band centered at 50 Hz is assumed, with an SNR of 10 dB at each

sensor, i.e.,  $G_{s,hh}(\omega)/G_w(\omega) = 10$  for  $h = 1, \dots, H$  and  $2\pi(49.5) < \omega < 2\pi(50.5)$  rad/sec. The signal coherence  $\gamma_{s,gh}(\omega) = \gamma_s(\omega)$  is varied between 0 and 1. We assume that  $T = 4000$  time samples are obtained at each sensor with sampling rate  $f_s = 2000$  samples/sec. The source localization performance is evaluated by computing the radius of the ellipse in  $(x, y)$  coordinates that satisfies the expression

$$\begin{bmatrix} x & y \end{bmatrix} \mathbf{J} \begin{bmatrix} x \\ y \end{bmatrix} = 1, \text{ where } \mathbf{J} \text{ is the FIM. If the errors}$$

in  $(x, y)$  localization are jointly Gaussian distributed, then the ellipse represents the contour at one standard deviation in root-mean-square (RMS) error. The error ellipse for any unbiased estimator of source location cannot be smaller than this ellipse derived from the FIM.

The  $H = 3$  arrays are located at coordinates  $(x_1, y_1) = (0, 0)$ ,  $(x_2, y_2) = (400, 400)$ ,  $(x_3, y_3) = (100, 0)$ , and one source is located at  $(x_s, y_s) = (200, 300)$ , where the units are meters. Figure 3a shows the ellipse radius for various values of the signal coherence  $\gamma_s(\omega)$ . Note that a significant improvement in localization accuracy is potentially possible with the small value of coherence  $\gamma_s(\omega) = 0.1$ , and the CRB gets smaller as the coherence increases. Note also that the localization scheme 2 described above (bearing plus time-delay estimation) may perform as well as the optimum, joint processing scheme.

The CRB results in Figure 3a indicate that even small amounts of signal coherence between widely distributed arrays provide the potential for significant improvement in source localization accuracy. We point out that the CRB results for time-delay estimation in this case are optimistic due to the narrowband model for the observations. With narrowband signals at 50 Hz, the time delays are resolvable only within the interval of one period of  $(50 \text{ Hz})^{-1} = 0.02$  sec. The CRB assumes that the ambiguities on the order of 0.02 seconds are resolved by an unbiased estimator. This ambiguity in time-delay estimation can be reduced by exploiting the wideband nature of the signals.

Next we present results from measured aeroacoustic data to illustrate typical values of signal coherence at distributed arrays. The experimental setup is illustrated in Figure 3b, which shows the path of a moving ground vehicle and the locations of four microphone arrays (labeled 1, 3, 4, 5). Each array is circular with  $N = 7$  sensors, 4-ft radius, and six sensors equally spaced around the perimeter with one sensor in the center. We focus on the 10 second segment indicated by the  $\diamond$ 's in Figure 3b. Figure 3c shows the power spectral density (PSD) of the data measured at arrays 1 and 3 during the 10 second segment. Note the dominant harmonic at 40 Hz. Figure 3d shows the estimated coherence between arrays 1 and 3 during the 10 second segment. The coherence is approximately 0.85 at 40 Hz, which demonstrates the presence of significant coherence at widely-separated microphones. Exploiting this coherence has the potential for improved source localization accuracy, as illustrated in the CRBs of Figure 3a. The Doppler effect due to source motion was compensated prior to the coherence estimate shown in Figure 3d.

## 5. REFERENCES

- [1] T. Pham and B. M. Sadler, "Adaptive wideband aeroacoustic array processing," *8th IEEE Statistical Signal and Array Processing Workshop*, pp. 295-298, Corfu, Greece, June 1996.
- [2] A. Paulraj and T. Kailath, "Direction of arrival estimation by eigenstructure methods with imperfect spatial coherence of wavefronts," *J. Acoust. Soc. Am.*, vol. 83, pp. 1034-1040, March 1988.
- [3] A.B. Gershman, C.F. Mecklenbrauker, J.F. Bohme, "Matrix fitting approach to direction of arrival estimation with imperfect spatial coherence," *IEEE Trans. on Signal Proc.*, vol. 45, no. 7, pp. 1894-1899, July 1997.
- [4] B.-G. Song and J.A. Ritcey, "Angle of arrival estimation of plane waves propagating in random media," *J. Acoust. Soc. Am.*, vol. 99, no. 3, pp. 1370-1379, March 1996.
- [5] D.K. Wilson, "Performance bounds for acoustic direction-of-arrival arrays operating in atmospheric turbulence," *J. Acoust. Soc. Am.*, vol. 103, no. 3, pp. 1306-1319, March 1998.
- [6] M. Wax and T. Kailath, "Decentralized processing in sensor arrays," *IEEE Trans. on Acoustics, Speech, Signal Processing*, vol. ASSP-33, no. 4, pp. 1123-1129, October 1985.
- [7] E. Weinstein, "Decentralization of the Gaussian maximum likelihood estimator and its applications to passive array processing," *IEEE Trans. Acoust., Speech, Sig. Proc.*, vol. ASSP-29, no. 5, pp. 945-951, October 1981.
- [8] P. Stoica, A. Nehorai, and T. Soderstrom, "Decentralized array processing using the MODE algorithm," *Circuits, Systems, and Signal Processing*, vol. 14, no. 1, 1995, pp. 17-38.
- [9] D.K. Wilson, "Atmospheric effects on acoustic arrays: a broad perspective from models," *1999 Meeting of the IRIS Specialty Group on Battlefield Acoustics and Seismics*, Laurel, MD, September 13-15, 1999.
- [10] B. Friedlander, "On the Cramer-Rao Bound for Time Delay and Doppler Estimation," *IEEE Trans. on Info. Theory*, vol. IT-30, no. 3, pp. 575-580, May 1984.
- [11] G.C. Carter (ed.), *Coherence and Time Delay Estimation* (Selected Reprint Volume), IEEE Press, 1993.
- [12] R.J. Kozick and B.M. Sadler, "Source localization with distributed sensor arrays and partial spatial coherence," *SPIE 2000 AeroSense Symp.*, Orlando, FL, April 24-28, 2000.

$$\mathbf{G}_Z(\omega) = \begin{bmatrix} \mathbf{a}_1(\omega)\mathbf{a}_1(\omega)^H G_{s,11}(\omega) & \cdots & \mathbf{a}_1(\omega)\mathbf{a}_H(\omega)^H \exp(-j\omega D_{1H})G_{s,1H}(\omega) \\ \vdots & \ddots & \vdots \\ \mathbf{a}_H(\omega)\mathbf{a}_1(\omega)^H \exp(+j\omega D_{1H})G_{s,1H}(\omega)^* & \cdots & \mathbf{a}_H(\omega)\mathbf{a}_H(\omega)^H G_{s,HH}(\omega) \end{bmatrix} + G_w(\omega)\mathbf{I} \quad (11)$$

$$J_{ij} = \frac{T}{2\omega_s} \int_0^{\omega_s} \text{trace} \left\{ \frac{\partial \mathbf{G}_Z(\omega)}{\partial \theta_i} \mathbf{G}_Z(\omega)^{-1} \frac{\partial \mathbf{G}_Z(\omega)}{\partial \theta_j} \mathbf{G}_Z(\omega)^{-1} \right\} d\omega, \quad i, j = 1, 2 \quad (12)$$

Figure 2: Cross-spectral density matrix  $\mathbf{G}_Z(\omega)$  of  $\mathbf{Z}(t)$  in (9), and FIM  $J$  for parameters  $\Theta = [x_s, y_s]^T$ .

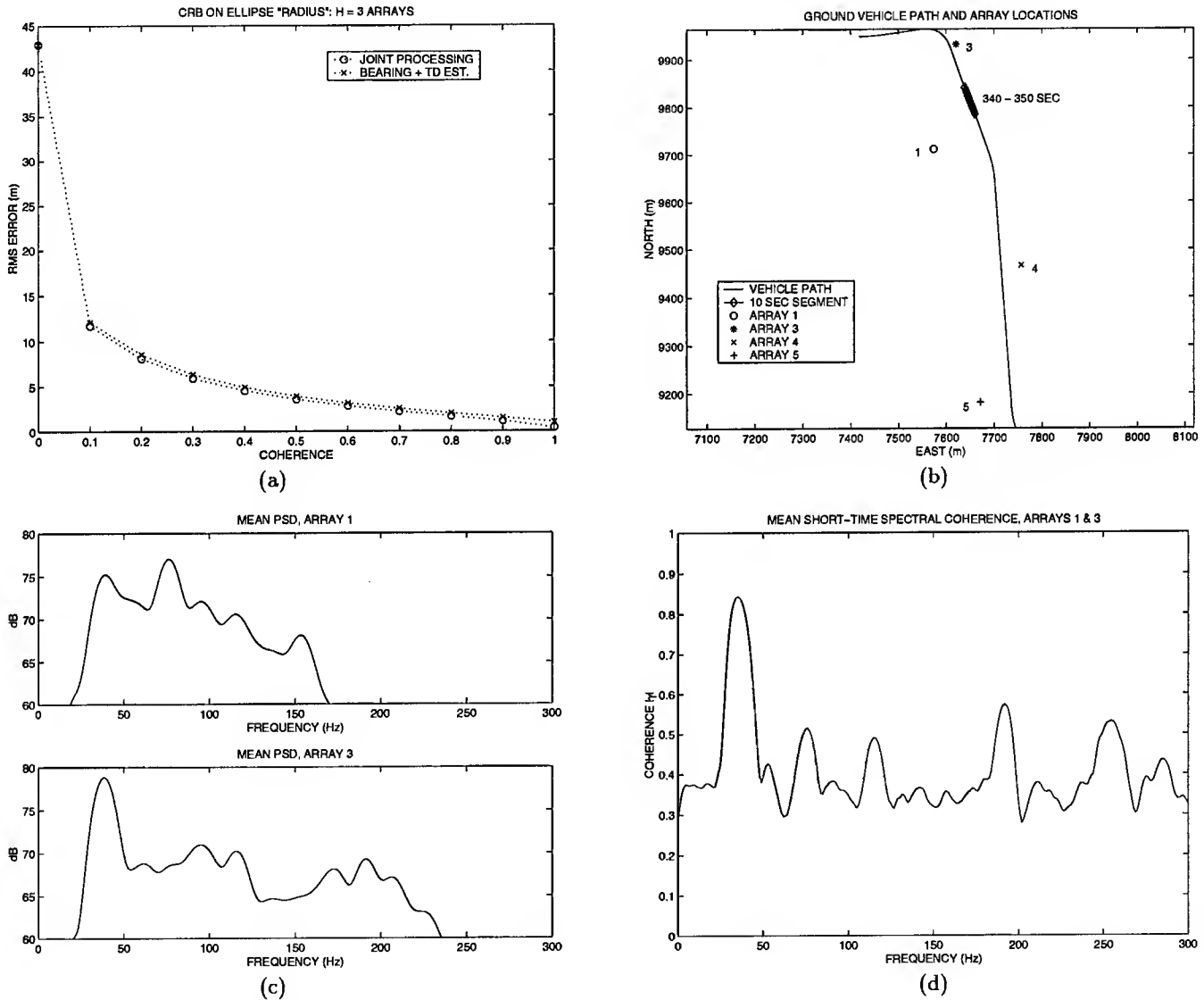


Figure 3: (a) CRBs on RMS source localization error for a scenario with  $H = 3$  arrays and one source. (b) Path of ground vehicle and array locations for measured data. (c) Mean power spectral density (PSD) at arrays 1 and 3 estimated from measured data over the 10 second segment  $\diamond$  in (b). Top panel is  $G_{s,11}(f)$ , bottom panel is  $G_{s,33}(f)$ . (d) Mean spectral coherence  $\gamma_{s,13}(f)$  estimated over the 10 second segment.

# DETERMINISTIC MAXIMUM LIKELIHOOD DOA ESTIMATION IN HETEROGENEOUS PROPAGATION MEDIA

*Petre Stoica<sup>\*</sup>, Olivier Besson<sup>†</sup> and Alex B. Gershman<sup>‡</sup>*

<sup>\*</sup> Department of Systems and Control, Uppsala University, Uppsala, Sweden

<sup>†</sup> Department of Avionics and Systems, ENSICA, Toulouse, France

<sup>‡</sup>Department of Electrical and Computer Engineering, McMaster University, Hamilton, Canada

## ABSTRACT

In a number of array signal processing applications, such as underwater source localization, the propagation medium is not homogeneous, which causes a distortion of the wavefront received by the array. In this paper, we consider the direction-of-arrival (DOA) estimation problem for such distorted wavefronts. In previous approaches, the so-called multiplicative noise scenario is considered based on the assumption that the distortion is random and can be parameterized by a small number of parameters. To gain robustness against mis-modelling we assume a scenario in which the wavefront amplitude is distorted in a completely arbitrary way. We derive the maximum likelihood (ML) estimator of the DOA and show it can be obtained by means of a simple 1-D search. The Cramér-Rao bound (CRB) for the problem at hand is derived. Numerical simulations illustrate a good performance of the estimator and show that its accuracy is comparable with that of estimators which require knowledge of the form of amplitude distortions.

## 1. INTRODUCTION AND PROBLEM FORMULATION

In a number of direction-of-arrival (DOA) estimation applications, such as underwater source localization by means of large hydrophone arrays, the heterogeneity of the propagation medium causes a distortion of the wavefront received by the array [1, 2]. More exactly, if we let  $y_k(t)$  denote the  $t$ -th observed sample of the output of the  $k$ -th sensor in the array and assume that the (distorted) wavefront impinging on the array is narrowband and its DOA is equal to  $\theta$ , then we can write,

for  $k = 1, \dots, m$  and  $t = 1, \dots, N$

$$y_k(t) = e^{i\phi(t)} x_k(t) a_k(\theta) + e_k(t) \quad (1)$$

where  $\phi(t) \in [-\pi, \pi]$  is the unknown time-varying phase of the received (baseband) wavefront at the first sensor,  $x_k(t) \in \mathbb{R}$  describes a time-and-space-varying amplitude distortion,  $e_k(t)$  is a noise term, and  $a_k(\theta) = e^{i\tau_k(\theta)}$  with  $\tau_k(\theta)$  being the DOA-dependent time needed by the wavefront to travel from the first to the  $k$ -th sensor (hence  $\tau_1(\theta) \equiv 0$ ). We assume that the noise  $\{e_k(t)\}$  is both spatially (in  $k$ ) and temporally (in  $t$ ) white and that it has a circular Gaussian distribution with zero mean and unknown variance  $\sigma^2$ . The problem we consider in the sequel is the maximum likelihood (ML) estimation of  $\theta$  (which is the parameter of interest) as well as  $\{\phi(t)\}_{t=1}^N$  and  $\{[x_k(t)]_{k=1}^m\}_{t=1}^N$  (which are the nuisance parameters) from the observed array data  $\{y_k(t)\}_{k=1, \dots, m}^{t=1, \dots, N}$ .

Most previous approaches to DOA estimation of distorted wavefronts have considered the so-called multiplicative noise scenario in which  $\{x_k(t)\}_{k=1}^m$  is assumed to be a spatially stationary, temporally white Gaussian random vector. While this assumption makes it possible to parameterize the distribution of  $\{x_k(t)\}$  by  $m$  parameters only (the spatial covariance matrix of  $\{x_k(t)\}_{k=1}^m$  is Toeplitz in such a case), it is evidently rather restrictive. In fact, to reduce the complexity of the DOA estimation problem even further, it is often assumed that the covariance matrix can be characterized by two parameters only. Even so, the (stochastic) ML estimation of those two parameters along with  $\theta$  and  $\sigma^2$  remains a complicated task requiring a computationally burdensome 4-D search [3]. Here, we take a different route, as suggested by the different set of assumptions that we have already made on (1). To make our DOA estimation approach robust to mis-modelling the wavefront amplitude distortion, we have modeled  $\{x_k(t)\}$  as arbitrary deterministic variables. This is a sensible thing to do when no *a priori* knowledge on

The work of P. Stoica was partly supported by the Senior Individual Grant Program of the Swedish Foundation for Strategic Research (SSF).

The work of A.B. Gershman was supported by the Natural Sciences and Engineering Research Council (NSERC) of Canada.

$\{x_k(t)\}$  is available, or when the observations contain a few samples only.

However, the (deterministic) ML estimation problem corresponding to the above assumption appears to be extremely complicated: besides the parameter of interest  $\theta$ , the likelihood function depends on  $(m+1)N+1$  nuisance parameters. Despite this apparent complexity, we show in the paper that all nuisance parameters can be eliminated from the likelihood function, hence leaving a 1-D search problem for the exact (deterministic) ML DOA estimation. Additionally, the so-obtained DOA estimate is quite accurate in spite of the large number of nuisance parameters present. In fact, we show that its variance is close to the Cramér-Rao bound (CRB), for which we derive a closed-form expression. Moreover, its accuracy compares favorably with that of the approximate (stochastic) ML DOA estimator of [3] obtained under the assumption (exploited by [3] but not by our DOA estimator) that the form of the covariance matrix of  $\mathbf{x}(t)$  is known. Finally, its robustness is evaluated. It is shown to perform reasonably well even in the presence of phase fluctuations of the wavefronts, i.e. when  $x_k(t)$  is complex-valued, even though its formulation ignores phase fluctuations and hence it is not intended for handling such a situation.

## 2. ML DOA ESTIMATION

Under the assumptions made the negative log-likelihood function can be written as [4, 5]

$$L = mN \log \pi + mN \log \sigma^2 + \frac{1}{\sigma^2} \sum_{t=1}^N \sum_{k=1}^m \left| y_k(t) - e^{i\phi(t)} x_k(t) a_k(\theta) \right|^2$$

It is well-known [4, 5] that  $L$  can be minimized explicitly with respect to  $\sigma^2$ , leaving a concentrated negative log-likelihood function that depends only on  $\theta$ ,  $\{x_k(t)\}$  and  $\{\phi(t)\}$ :

$$\tilde{L} = \sum_{t=1}^N \sum_{k=1}^m \left| y_k(t) - e^{i\phi(t)} x_k(t) a_k(\theta) \right|^2 \quad (2)$$

The minimization of the function above with respect to  $\{\phi(t)\}$  and  $\{x_k(t)\}$  reduces to the minimization of the inner term sum in (2) for each  $t$ . Therefore, let us consider the following generic function (where the dependence on  $t$  and  $\theta$  is temporarily omitted for notational convenience):

$$f = \sum_{k=1}^m \left| y_k - e^{i\phi} x_k a_k \right|^2 \quad (3)$$

A straightforward calculation shows that

$$f = \sum_{k=1}^m |y_k|^2 + [x_k - \text{Re}(e^{-i\phi} a_k^* y_k)]^2 - [\text{Re}(e^{-i\phi} a_k^* y_k)]^2 \quad (4)$$

where the superscript  $*$  denotes the complex conjugate for scalars and the conjugate transpose for matrices and vectors. The minimization of (4) with respect to  $x_k(t)$  yields:

$$\hat{x}_k(t) = \text{Re} \left[ e^{-i\hat{\phi}(t)} a_k^*(\hat{\theta}) y_k(t) \right] \quad (5)$$

where the ML estimators (MLE's)  $\hat{\theta}$  and  $\{\hat{\phi}(t)\}$  are yet to be determined. Insertion of (5) into (4) shows that the MLE of  $\phi(t)$  is obtained by maximizing (for each  $t$ ) the function

$$\begin{aligned} g &= 2 \sum_{k=1}^m [\text{Re}(e^{-i\phi} a_k^* y_k)]^2 \\ &= \sum_{k=1}^m \left[ |y_k|^2 + \text{Re}(e^{-i2\phi} a_k^{2*} y_k^2) \right] \\ &= \text{const.} + \left| \sum_{k=1}^m a_k^{2*} y_k^2 \right| \cos \left[ \arg \left( \sum_{k=1}^m a_k^{2*} y_k^2 \right) - 2\phi \right] \end{aligned} \quad (6)$$

where, to derive the second equality, we used the fact that for any complex number  $\alpha$ ,

$$[\text{Re}(\alpha)]^2 = \frac{1}{4} (\alpha + \alpha^*)^2 = \frac{1}{2} \left[ |\alpha|^2 + \text{Re}(\alpha^2) \right]$$

It follows that

$$\hat{\phi}(t) = \frac{1}{2} \arg \left( \sum_{k=1}^m a_k^{2*}(\hat{\theta}) y_k^2(t) \right) \quad (7)$$

with the MLE of  $\theta$  given by

$$\hat{\theta} = \arg \max_{\theta} \sum_{t=1}^N \left| \sum_{k=1}^m a_k^{2*}(\theta) y_k^2(t) \right| \quad (8)$$

Hence, the deterministic ML DOA estimator simply entails a 1-D search. For uniform linear arrays (ULA), we have  $a_k(\theta) = \exp\{i(k-1)\omega\}$  where  $\omega = 2\pi\Delta \sin \theta$  is the so-called spatial frequency and  $\Delta$  is the inter-element spacing in wavelengths. In such a case, the inner sum in (8) can be evaluated (as a function of  $\theta$ ) by using a FFT algorithm with zero padding applied to the squared data samples  $\{y_k^2(t)\}_{k=1}^m$  (for each  $t$ ). The ability to do so is particularly valuable, from a computational standpoint, for large arrays ( $m \gg 1$ ).

**Remark 1** Note that if we re-define the data samples as  $z_k(t) = y_k^2(t)$ , and the steering vector elements as  $\alpha_k(\theta) = a_k^2(\theta)$  then  $\hat{\theta} = \arg \min_{\theta} \sum_{t=1}^N |\alpha^*(\theta) z(t)|$ . The function in (8) can thus be interpreted as an  $L_1$ -beamformer, except for the squaring operation applied in (8) to the observed data and the elements of the steering vector prior to beamforming.

### 3. CRAMÉR-RAO BOUNDS

In this section, we derive the Cramér-Rao Bounds for the problem at hand. For the sake of clarity, let us introduce the following notations

$$\begin{aligned} \mathbf{a}(\theta) &= [e^{i\tau_1(\theta)}, e^{i\tau_2(\theta)}, \dots, e^{i\tau_m(\theta)}]^T \\ \phi &= [\phi(1), \dots, \phi(N)]^T \\ \mathbf{x}(t) &= [x_1(t), \dots, x_m(t)]^T \\ \tilde{\mathbf{x}} &= [\mathbf{x}^T(1), \dots, \mathbf{x}^T(N)]^T \end{aligned}$$

It is well-known [4, 5] that under the assumptions made the CRB for the noise variance is decoupled from the CRB for the other parameters. In the following, we concentrate on the CRB for  $\eta = [\tilde{\mathbf{x}}^T \ \phi^T \ \theta]^T$ . The Fisher Information Matrix (FIM) is given by [4, 5]

$$\mathbf{F}(k, \ell) = \frac{2}{\sigma^2} \text{Re} \left[ \sum_{t=1}^N \frac{\partial \mu^*(t)}{\partial \eta_k} \frac{\partial \mu(t)}{\partial \eta_\ell} \right] \quad (9)$$

with

$$\mu(t) = \mathcal{E} \{y(t)\} = e^{i\phi(t)} \Phi_a(\theta) \mathbf{x}(t) \quad (10)$$

and where  $\Phi_a(\theta) = \text{diag}(\mathbf{a}(\theta))$ . In order to derive a closed-form expression for the FIM, first note that

$$\frac{\partial \mu(t)}{\partial x_k(s)} = e^{i\phi(t)} \Phi_a(\theta) \mathbf{e}_k \delta(t, s) \quad (11)$$

$$\frac{\partial \mu(t)}{\partial \phi(s)} = i e^{i\phi(t)} \Phi_a(\theta) \mathbf{x}(t) \delta(t, s) \quad (12)$$

$$\frac{\partial \mu(t)}{\partial \theta} = e^{i\phi(t)} \Phi_d(\theta) \mathbf{x}(t) \quad (13)$$

where  $\Phi_d(\theta) = \text{diag}(\mathbf{d}(\theta))$  and  $\mathbf{d}(\theta) = \partial \mathbf{a}(\theta) / \partial \theta$ .  $\delta(t, s)$  is the Kronecker delta, and  $\mathbf{e}_k$  is the  $m$ -dimensional vector with all elements equal to zero, except the  $k$ -th element which equals one. Using the previous results, it can be shown (see [6] for details) that the FIM has the following block form

$$\mathbf{F} = \frac{2}{\sigma^2} \begin{bmatrix} \mathbf{I}_{mN} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{F}_\phi & \mathbf{F}_{\phi\theta} \\ \mathbf{0} & \mathbf{F}_{\phi\theta}^T & \mathbf{F}_\theta \end{bmatrix} \quad (14)$$

where  $(t, r = 1, \dots, N)$

$$\mathbf{F}_\phi(t, r) = \|\mathbf{x}(t)\|^2 \delta(t, r) \quad (15)$$

$$\mathbf{F}_\theta = \sum_{t=1}^N \mathbf{x}^T(t) \mathbf{T}^2 \mathbf{x}(t) \quad (16)$$

$$\mathbf{F}_{\phi\theta}(r) = \mathbf{x}^T(r) \mathbf{T} \mathbf{x}(r) \quad (17)$$

and  $\mathbf{T} \triangleq \text{diag}(\tau'_1(\theta), \tau'_2(\theta), \dots, \tau'_m(\theta))$ . The block corresponding to  $[\phi^T \ \theta]^T$  is

$$\bar{\mathbf{F}} \triangleq \frac{2}{\sigma^2} \begin{bmatrix} \mathbf{F}_\phi & \mathbf{F}_{\phi\theta} \\ \mathbf{F}_{\phi\theta}^T & \mathbf{F}_\theta \end{bmatrix}$$

The CRB for  $\theta$  is obtained as the lower-right corner element of  $\bar{\mathbf{F}}^{-1}$ . By a formula for the inverse of partitioned matrices [4, 5], it can be readily shown that the CRB for  $\theta$  can be written as follows

$$\begin{aligned} \text{CRB}(\theta) &= \frac{\sigma^2}{2} \frac{1}{\mathbf{F}_\theta - \mathbf{F}_{\phi\theta}^T \mathbf{F}_\phi^{-1} \mathbf{F}_{\phi\theta}} \\ &= \frac{\sigma^2}{2} \left\{ \sum_{t=1}^N \left[ \mathbf{x}^T(t) \mathbf{T}^2 \mathbf{x}(t) - \frac{[\mathbf{x}^T(t) \mathbf{T} \mathbf{x}(t)]^2}{\mathbf{x}^T(t) \mathbf{x}(t)} \right] \right\}^{-1} \end{aligned} \quad (18)$$

### 4. NUMERICAL EXAMPLES AND CONCLUSIONS

In this section, we illustrate the performance of our deterministic MLE and compare it with the COMET estimator [3], which is a large-sample realization of the stochastic ML estimator. A comparison with the  $L_1$ -beamformer

$$\hat{\theta}^1 = \arg \max_{\theta} \sum_{t=1}^N \left| \sum_{k=1}^m a_k^*(\theta) y_k(t) \right|$$

and the conventional  $L_2$ -beamformer

$$\hat{\theta}^2 = \arg \max_{\theta} \sum_{t=1}^N \left| \sum_{k=1}^m a_k^*(\theta) y_k(t) \right|^2$$

is also presented. We consider a ULA of  $m = 16$  sensors spaced a half wavelength apart. The DOA of the source is set to  $\theta = 10^\circ$ .  $\mathbf{x}(t)$  is a real-valued, zero-mean temporally white Gaussian random process with covariance matrix  $\mathbf{R} = \mathcal{E} \{ \mathbf{x}(t) \mathbf{x}^T(t) \}$ . To make COMET applicable, we assume that the elements of  $\mathbf{R}$  are given by  $\mathbf{R}(k, \ell) = \rho^{|k-\ell|}$  with  $\rho = 0.9$  which corresponds to a  $10 \log_{10} \rho^2 = -0.915$  dB coherence loss at one wavelength separation [2]. The signal to noise ratio (SNR) is defined as  $-10 \log_{10}(\sigma^2)$ . 300 Monte-Carlo simulations were run to estimate the root mean-square error

(RMSE) of the estimates, with all values given in degrees ( $^\circ$ ). For each of the 300 simulations, a different realization of  $\{x(t)\}_{t=1}^N$  was used and the corresponding (deterministic) CRB was computed from (18). The so-obtained values were then averaged over the 300 trials to yield what we refer to as the average deterministic CRB. Finally, we also display the stochastic CRB derived under the assumption that  $x(t)$  is a white Gaussian random process with a covariance matrix  $R$  parameterized by  $\rho$ .

Fig. 1 displays the RMSE of the estimates versus the number of snapshots for  $SNR = 0$  dB. It can be observed that the deterministic MLE has a performance close to the deterministic CRB. In all cases, but especially in low samples, the deterministic ML estimator outperforms the COMET estimator in spite of the fact that the former is computationally simpler and does not require as many assumptions as COMET does. The deterministic MLE performs better than the CBF, especially in small samples, and always outperforms the  $L_1$ -beamformer significantly.

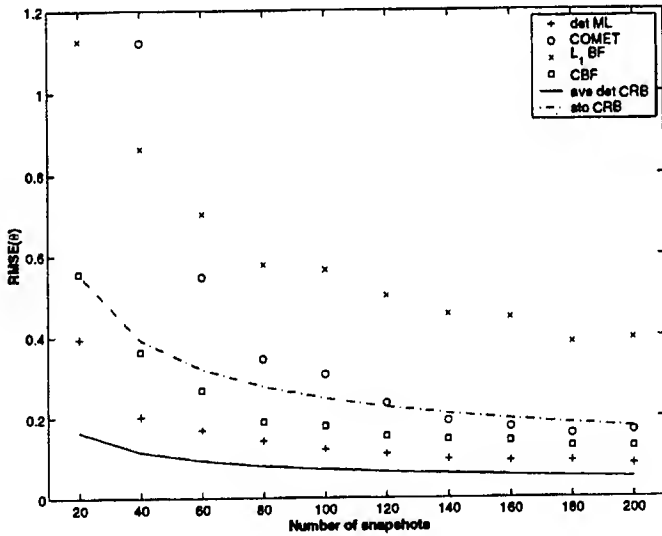


Figure 1: CRB's and RMSE of the estimators  $N$ . Amplitude distortions.  $SNR = 0$  dB.

Next, the influence of the number of sensors on the estimation performance is examined in Fig. 2 where  $m$  is varied from 8 to 40 while the number of snapshots is fixed at  $N = 60$  and  $SNR = 0$  dB. Fig. 2 reveals that the empirical RMSE of the deterministic ML estimator remains nearly constant while that of the other estimators, particularly COMET, increases when  $m$  increases. Finally, we study the influence of  $\rho$  on the performance of the estimators. The coherence loss at a wavelength separation is varied from  $-3$  to  $-0.25$  dB while  $N = 60$  and  $SNR = 0$  dB. The re-

sults are plotted in Fig. 3. It can be noticed that the performance of the deterministic MLE remains nearly constant for coherence losses less than  $-1.5$  dB, and tends to increase under this value. We have observed that this threshold depends on  $m$  and decreases when  $m$  decreases, which seems logical. On the other hand, all other estimators have a performance that degrades continuously with the coherence loss. Finally, we note that the CBF is as accurate as the deterministic ML for small coherence loss values.

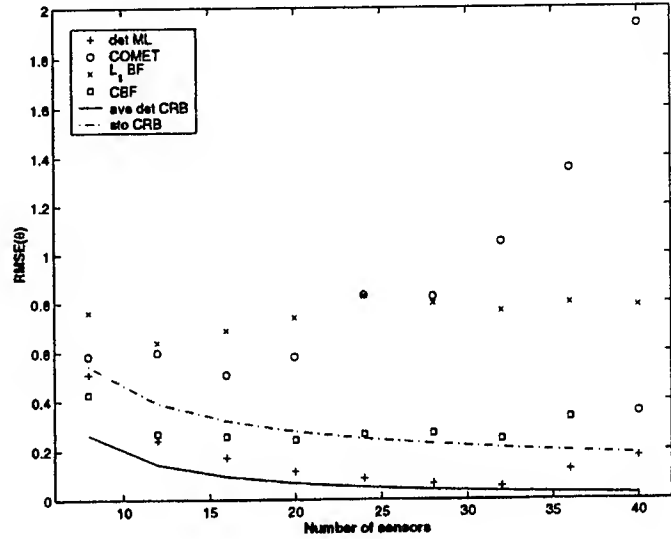


Figure 2: CRB's and RMSE of the estimators versus the number of sensors. Amplitude distortions.  $N = 60$  and  $SNR = 0$  dB.

In a second series of simulations, we test the robustness of our estimator. We consider the more complicated situation where the distortions affect not only the wavefront amplitude but the phase as well. To this end, phase fluctuations are introduced in the model. More exactly,  $x(t)$  is modeled as  $x_k(t) = \tilde{x}_k(t)e^{i\psi_k(t)}$ , where  $\tilde{x}(t)$  is a real-valued process with the same covariance matrix  $R$  as before and  $\psi(t)$  is a zero-mean sensor-to-sensor independent-increment process, i.e.  $\psi_k(t) = \psi_{k-1}(t) + \Delta\psi_k(t)$  where  $\Delta\psi_k$  are independent random variables uniformly distributed on  $[-\pi\delta, \pi\delta]$ . The variance of the phase distortions is  $\mathcal{E}\{\psi_k^2\} = (k-1)\delta^2\pi^2/3$  and hence increases with the sensor index. Note that the previous set of simulations considered the case of  $\delta = 0$  and also observe that the deterministic MLE is not intended for handling the case of  $\delta \neq 0$ . In contrast, the COMET estimator can cope with this problem since it only relies on the form of the covariance matrix of the observations, which is unchanged. In other words, the deterministic MLE assumes no phase distortions and hence uses a *wrong* model whereas COMET utilizes a correct model. This example is chosen to test

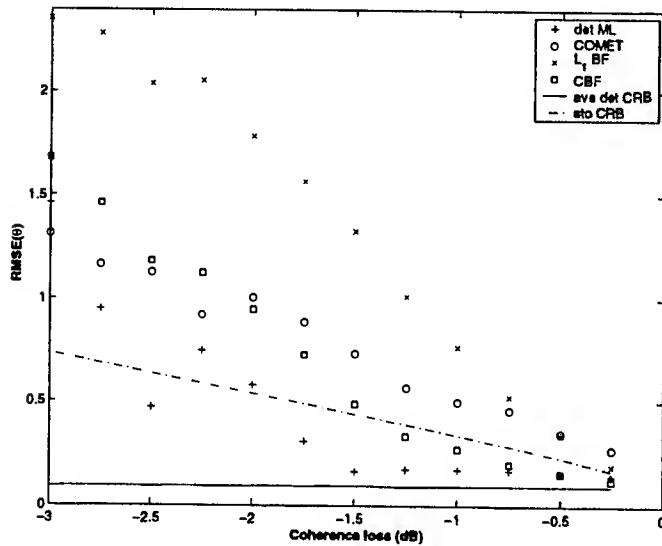


Figure 3: CRB's and RMSE of the estimators versus the coherence loss. Amplitude distortions.  $N = 60$  and  $SNR = 0$  dB.

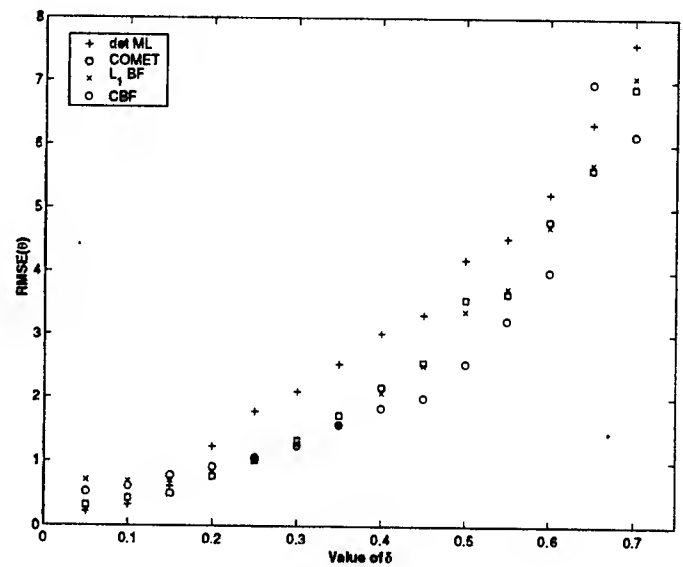


Figure 4: RMSE of the estimators versus  $\delta$ . Amplitude and phase distortions.  $N = 60$  and  $SNR = 0$  dB.

the robustness of our estimator and check whether it can be applied in more difficult scenarios. Evidently, the CRB formula (18) is no longer relevant. Moreover, the stochastic CRB which has been derived under the Gaussian assumption can no longer be used. Fig. 4 displays the RMSE of the estimates when  $\delta$  is varied. We consider a scenario with a relatively small number of snapshots  $N = 60$  and a low  $SNR = 0$  dB. It can be noted that the performance of the deterministic ML estimator as well as that of the other estimators continuously degrades as  $\delta$  increases. For  $\delta$  above a threshold, COMET and the CBF have a smaller RMSE, the former having the smallest RMSE. However, the deterministic MLE proposed herein turns out to be quite robust and accurate for a wide range of phase fluctuations, which is an additional interesting feature of it.

The deterministic MLE proposed in this paper does not make any modeling assumption on the amplitude distortion, which is a significant advantage. On the other hand, it assumes that there is no phase distortion (so that  $x(t)$  is real-valued). We showed that it is quite robust to the violation of this latter assumption, yet its performance does degrade as the phase distortion increases. The trade-off between the stochastic MLE (for which COMET is an approximate (large-sample) implementation) and the deterministic MLE of this paper is hence quite clear: depending on the *a priori* information available on the wavefront distortion, one approach may be preferred to another, with the deterministic MLE having more chances to be chosen in scenarios with little or no *a priori* information.

## REFERENCES

- [1] A. Paulraj and T. Kailath, "Direction of arrival estimation by eigenstructure methods with imperfect spatial coherence of wavefronts," *Journal Acoust. Society Amer.*, vol. 83, pp. 1034–1040, March 1988.
- [2] A. Gershman, C. Mecklenbräuker, and J. Böhme, "Matrix fitting approach to direction of arrival estimation with imperfect coherence of wavefronts," *IEEE Transactions Signal Processing*, vol. 45, pp. 1894–1899, July 1997.
- [3] T. Trump and B. Ottersten, "Estimation of nominal direction of arrival and angular spread using an array of sensors," *Signal Processing*, vol. 50, pp. 57–70, April 1996.
- [4] S. Kay, *Fundamentals of Statistical Signal Processing: Estimation Theory*. Englewood Cliffs, NJ: Prentice Hall, 1993.
- [5] P. Stoica and R. Moses, *Introduction to Spectral Analysis*. Upper Saddle River, NJ: Prentice Hall, 1997.
- [6] P. Stoica, O. Besson, and A. Gershman, "Array processing for heterogeneous propagation media." submitted for publication, 2000.



# EFFICIENT SIGNAL DETECTION IN PERTURBED ARRAYS

Anil M. Rao and Douglas L. Jones

Coordinated Science Laboratory  
University of Illinois  
1308 W. Main Street  
Urbana, IL 61801

anilrao@dsp.csl.uiuc.edu, tel: (217) 333-5860

jones@dsp.csl.uiuc.edu, tel: (217) 244-6823

Fax: (217) 244-1642

## ABSTRACT

The use of sensor arrays in signal processing applications has received considerable attention. Various array perturbations caused by phase, calibration, or modeling errors often cause the sensor observations to become only partially correlated, limiting the application of traditional matched-field beamformers. Quadratic array processing is optimal for many randomly perturbed array problems; however, direct implementation poses a significant computational burden. We propose a highly efficient, asymptotically optimal method of implementing quadratic array processors suitable for detection problems in randomly perturbed arrays. Specifically, we show that under certain conditions the optimal array processor can be approximately realized efficiently and robustly employing only discrete Fourier transforms to deal with spatial processing while entailing only a small loss in performance.

## 1. INTRODUCTION

The detection of signals in noise is a classical hypothesis testing problem. The use of a sensor array can considerably enhance signal detection by providing a large gain in the SNR and allowing for target or signal source localization. Due to the need for fast processing of target data in radar/sonar applications, the need exists for very efficient array processing structures. Unfortunately, perturbations in the array or imperfect spatial coherence of the signal wavefronts due to complicated propagation may lead to complex receiver structures for optimal performance. The assumptions of a known array response is rarely satisfied in practice. Due to changes in the weather, the surrounding environment, and antenna location, the response of the array may be significantly different than when it was last calibrated [1]. If the perturbations in the array are deterministic and known, then it is easy to compensate for them and traditional matched-field beamforming detector structures (in which the observations are aligned, summed together, and correlated against the signal of interest) can still be used [2, 3, 4]. However, array perturbations are most often unknown and must be modeled as random, leading to optimal detection structures which

are quadratic in the observations; such detectors require complicated matrix combining in space. While it is relatively straightforward to derive the form of the optimal quadratic processor, its implementation is computationally too expensive for many real-time applications. For this reason, most literature regarding random perturbations in arrays has focused only on the effects of such perturbations on the performance of traditional processors which ignore the perturbations; the effect of phase errors is studied in [5], the effect of spatial calibration errors on detection performance is studied in [6], the effect of model errors on the performance of the MUSIC algorithm is studied in [1, 7], and loss of signal coherence as it propagates across the array is studied in [2]. The results of these analyses demonstrate that random perturbations can cause significant degradation in performance of traditional processors.

In this paper we propose a highly efficient technique of implementing an asymptotically optimal detector for dealing with array perturbations. We develop a novel method of employing the popular Fourier transform algorithm to deal with the quadratic nature of the detection problem; specifically, we show that under certain conditions the discrete Fourier transform (DFT) is asymptotically optimal for spatial processing. Furthermore, we show that conventional frequency domain techniques for angle of arrival searches can easily be incorporated into our framework. In the end, we show that our proposed processor provides a significant improvement in performance over existing, traditional processors while achieving a comparable cost of implementation.

## 2. MATHEMATICAL PRELIMINARIES

Let us assume that we are dealing with an  $M$ -sensor uniform linear array with spacing  $d$ , and that a single  $N$ -sample signal  $\mathbf{s} = [s(1), \dots, s(N)]^T$  comes in to the array at angle  $\theta$ , where  $\theta$  is usually unknown. We will denote the  $N$ -sample observation at the  $i$ th sensor with the column vector  $\mathbf{x}_i = [x_i(1), \dots, x_i(N)]^T$ ,  $i = 1, \dots, M$ . Each sensor observation will have a component due to the signal  $\mathbf{s}$  and an additive noise component which we denote  $\mathbf{n}_i$ . Let  $\mathbf{X}^T = [\mathbf{x}_1^T, \dots, \mathbf{x}_M^T]^T$  denote the sensor observations concatenated into an  $MN \times 1$  column vector, and let  $\mathbf{N}$  be similarly defined. We assume the noise is white, Gaussian, uncorrelated between sensors, and independent of the signal source; we

This work was supported by the Office of Naval Research, contract no. N00014-95-1-0674

write its covariance matrix as  $E[\mathbf{N}\mathbf{N}^H] = \sigma^2\mathbf{I}$ . As in previous work dealing with array perturbations, we use narrowband array assumptions [4, 6]. A problem is classified as narrowband if the signal bandwidth is small compared to the inverse of the transit time of the wavefront across the array. This allows us to approximate the time delay that the signal encounters as it propagates between sensors as phase shifts; of course this would hold exactly if the signal source were sinusoidal over the block of samples collected. We may write the observation in Kronecker form as

$$\mathbf{X} = \mathbf{a}(\theta) \otimes \mathbf{s} + \mathbf{N} \quad (1)$$

where  $\mathbf{a}(\theta) = [a_1(\theta), a_2(\theta), \dots, a_M(\theta)]^T$  is the length- $M$  complex array response vector. The vector  $\mathbf{a}(\theta)$  contains all information as to how the signal component is related between sensors. Often  $\mathbf{a}(\theta)$  is taken to be a deterministic quantity; in a uniform linear array in which the signal has carrier frequency  $f_0$  and the speed of propagation is  $c$ , we have [4]

$$\mathbf{a}_m(\theta) = [1, e^{-j2\pi f_0 d \sin(\theta)/c}, \dots, e^{-j2\pi(M-1)f_0 d \sin(\theta)/c}]^T, \quad (2)$$

which represents a pure phase delay between each element. We will see later that due to the similarity with the DFT basis functions, such an array response vector allows for computationally efficient FFT algorithms to search over the unknown angle of arrival  $\theta$ . However, as discussed in the introduction, there are often random perturbations in the array. A very popular method of modeling these random perturbations is to assume the array response  $\mathbf{a}(\theta)$  is a random quantity. Here it is common to assume that  $\mathbf{a}(\theta)$  has a mean (nominal, calibrated) value  $\mathbf{a}_m(\theta)$ , such as in (2), plus a zero-mean random component  $\tilde{\mathbf{a}}(\theta)$ :

$$\mathbf{a}(\theta) = \mathbf{a}_m(\theta) + \tilde{\mathbf{a}}(\theta). \quad (3)$$

The random term  $\tilde{\mathbf{a}}(\theta)$  will have the form  $\mathbf{A}(\theta)[\tilde{a}_1, \dots, \tilde{a}_M]^T$  where  $\mathbf{A}(\theta)$  is a diagonal matrix containing the elements of  $\mathbf{a}_m(\theta)$  on the diagonal and the  $\tilde{a}_i$ 's are complex random quantities that represent gain and phase errors. Typically the vector  $[\tilde{a}_1, \dots, \tilde{a}_M]^T$  is assumed to have a Gaussian distribution with known covariance  $\mathbf{R}_{\tilde{\mathbf{a}}}$  [1, 6, 7, 8]. A diagonal  $\mathbf{R}_{\tilde{\mathbf{a}}}$  would imply the array errors are uncorrelated; off-diagonal terms would indicate sensor-to-sensor correlations that result if some sensors, such as adjacent elements, tend to perturb uniformly. The covariance of  $\tilde{\mathbf{a}}(\theta)$  is given by  $\mathbf{R} = \mathbf{A}(\theta)\mathbf{R}_{\tilde{\mathbf{a}}}\mathbf{A}^H(\theta)$ .

For clarity of presentation, we will pose our detection problem in the radar signal detection setting. We assume a known, deterministic signal  $\mathbf{s}$  is transmitted. If a target is present, the reflected signal is assumed to be  $b\mathbf{s}$ , where  $b$  is a deterministic but unknown complex phase factor. For simplicity we will assume that  $|b| \equiv 1$  so that the SNR of the reflected signal is known. When the signal is present, the observation is given by

$$\mathbf{X} = (\mathbf{a}_m(\theta) + \tilde{\mathbf{a}}(\theta)) \otimes b\mathbf{s} + \mathbf{N}, \quad (4)$$

where  $\tilde{\mathbf{a}}(\theta) \sim \mathcal{N}(0, \mathbf{R})$ . We write the net signal component as two terms  $\mathbf{S} = \mathbf{S}_m(b) + \tilde{\mathbf{S}}$  where  $\mathbf{S}_m(b) = \mathbf{a}_m(\theta) \otimes b\mathbf{s}$  is the signal mean and  $\tilde{\mathbf{S}} = \tilde{\mathbf{a}}(\theta) \otimes b\mathbf{s}$  is a zero-mean Gaussian component with covariance  $\mathbf{R}_{\tilde{\mathbf{S}}}$ . We assume that the noise is Gaussian and white in both time and space with variance  $\sigma^2$ .

### 3. OPTIMAL QUADRATIC ARRAY PROCESSOR

In this section we derive the form of the optimal processor and discuss its implementation. We will find that the optimal structure

is quadratic in the observations and that implementation requires decorrelating the observations in space. In statistical hypothesis testing, for an observation,  $\mathbf{X}$ , a real-valued test statistic,  $L(\mathbf{X})$ , is compared to a threshold to decide in favor of  $H_0$ , only noise is present, or  $H_1$ , the signal,  $\mathbf{S}$ , is present. The optimal test statistic based on the likelihood ratio is given by [9]:

$$L(\mathbf{X}) = \frac{1}{2} \log \frac{|\sigma^2\mathbf{I}|}{|\mathbf{R}_{\tilde{\mathbf{S}}} + \sigma^2\mathbf{I}|} + \frac{1}{2} \mathbf{X}^H (\sigma^2\mathbf{I} - (\mathbf{R}_{\tilde{\mathbf{S}}} + \sigma^2\mathbf{I})^{-1}) \mathbf{X} \\ + \text{Re} \left\{ \mathbf{X}^H (\mathbf{R}_{\tilde{\mathbf{S}}} + \sigma^2\mathbf{I})^{-1} \mathbf{S}_m(b) \right\} + \frac{1}{2} \mathbf{S}_m^H(b) (\mathbf{R}_{\tilde{\mathbf{S}}} + \sigma^2\mathbf{I})^{-1} \mathbf{S}_m(b).$$

It is common to use a generalized likelihood ratio test (GLRT) to deal with the unknown parameter  $b$ , which involves maximizing  $L(\mathbf{X})$  with respect to  $b$  and using that value of  $b$  which attains the maximum. The second and third terms in the expression for the likelihood ratio are quadratic and linear in the observations, respectively, while the first and fourth terms are just constants. It can be shown that after maximization over  $b$ , the optimal test statistic, retaining only those terms which depend on the observation, may be written as

$$L(\mathbf{X}) = \frac{1}{2\sigma^2} \mathbf{X}^H \mathbf{R}_{\tilde{\mathbf{S}}} (\mathbf{R}_{\tilde{\mathbf{S}}} + \sigma^2\mathbf{I})^{-1} \mathbf{X} + \left| \mathbf{S}_m^H (\mathbf{R}_{\tilde{\mathbf{S}}} + \sigma^2\mathbf{I})^{-1} \mathbf{X} \right| \\ = Q(\mathbf{X}) + T(\mathbf{X}),$$

where  $Q(\mathbf{X})$  denotes the quadratic term,  $T(\mathbf{X})$  denotes the linear term, and  $\mathbf{S}_m = \mathbf{a}(\theta) \otimes \mathbf{s}$ . We first discuss implementation of the quadratic term, which will involve an eigendecomposition. Because  $\tilde{\mathbf{S}} = \tilde{\mathbf{a}}(\theta) \otimes b\mathbf{s}$  and  $|b| = 1$ , its covariance  $\mathbf{R}_{\tilde{\mathbf{S}}}$  may be expressed as  $\mathbf{R}_{\tilde{\mathbf{S}}} = \mathbf{A}(\theta)\mathbf{R}_{\tilde{\mathbf{a}}}\mathbf{A}^H(\theta) \otimes \mathbf{s}\mathbf{s}^H$ . We may express the covariance matrix in eigenform as  $\mathbf{R}_{\tilde{\mathbf{S}}} = \mathbf{U}\mathbf{A}\mathbf{U}^H$ . The eigenvector matrix  $\mathbf{U}$  may be expressed with a Kronecker product as  $\mathbf{U} = \frac{\mathbf{s}}{\|\mathbf{s}\|} \otimes \mathbf{A}(\theta)\mathbf{U}_{\tilde{\mathbf{a}}}$ , where we assume  $\mathbf{U}_{\tilde{\mathbf{a}}}$ , the eigenvector matrix for  $\mathbf{R}_{\tilde{\mathbf{a}}}$ , is of full dimension  $M \times M$ . Let  $\mathbf{Z} = \mathbf{U}^H \mathbf{X} = \mathbf{U}_{\tilde{\mathbf{a}}}^H \mathbf{A}^H(\theta) (\frac{\mathbf{s}}{\|\mathbf{s}\|} \otimes \mathbf{X}) = [z_1, \dots, z_M]^T$  represent the spatially decorrelated, aligned matched filter samples (decorrelate via  $\mathbf{U}_{\tilde{\mathbf{a}}}^H$  and align via  $\mathbf{A}^H(\theta)$ ). It can be shown that the quadratic term may be implemented as

$$Q(\mathbf{X}) = \frac{1}{2\sigma^2} \sum_{k=1}^M \frac{\|\mathbf{s}\|^2 \lambda_k}{\|\mathbf{s}\|^2 \lambda_k + \sigma^2} |z_k|^2 \quad (5)$$

where the  $\lambda_k$ 's are the eigenvalues of  $\mathbf{R}_{\tilde{\mathbf{a}}}$ . This reveals that the quadratic term in the detector requires matched-filtering each sensor observation against the signal source in time, aligning the samples, spatially processing via the matrix  $\mathbf{U}_{\tilde{\mathbf{a}}}^H$  to obtain the decorrelated samples  $z_k$ ,  $k = 1, \dots, M$ , and then combining terms in the proper fashion. Hence decoupled spatial and temporal processing is optimal.

We would like to point out that the alignment via  $\mathbf{A}^H(\theta)$  may be carried out efficiently via Fourier transforms as in traditional narrowband array processing with no perturbations. If we let  $\mathbf{Y} = \frac{\mathbf{s}}{\|\mathbf{s}\|} \otimes \mathbf{X} = [y_1, \dots, y_M]^T$  denote the matched filter samples, then  $\mathbf{A}^H(\theta)\mathbf{Y} = [y_1, y_2 e^{j2\pi\phi}, \dots, y_M e^{j2\pi(M-1)\phi}]^T$ , where we have let  $\phi = -\frac{f_0 d}{c} \sin(\theta)$  represent the mapping from angle to spatial frequency. Therefore, when processing with  $\mathbf{U}_{\tilde{\mathbf{a}}}^H = \{u_{lk}\}$ , the  $l$ th element of the  $M \times 1$  column vector resulting from the matrix product

$\mathbf{Z} = \mathbf{U}_a^H \mathbf{A}^H(\theta) \mathbf{Y}$  is given by

$$\mathbf{Z}_l = \sum_{k=1}^M u_{lk} y_k e^{-j2\pi(k-1)\phi}. \quad (6)$$

Letting  $\tilde{y}_l(k) = M u_{lk} y_k$ , we may write the aligned, decorrelated matched filter samples as

$$\mathbf{Z} = [\tilde{Y}_1(\phi), \dots, \tilde{Y}_M(\phi)]^T \quad (7)$$

where  $\tilde{Y}_l(\phi)$  denotes the DTFT of the weighted observation snapshot  $\{\tilde{y}_l(k)\}_{k=1}^M$ . Hence each decorrelated aligned matched filter sample can be obtained as a function of the angle of arrival through a DTFT, allowing the use of computationally efficient FFT-based algorithms when we must search over an unknown  $\theta$  as in a GLRT.

We must be careful in how we choose to implement the linear term  $T(\mathbf{X})$  to insure that FFT-based algorithms can still be used to search over the unknown angle of arrival  $\theta$ . Using techniques similar to those used to write the quadratic term, the linear term may be written as

$$T(\mathbf{X}) = |\mathbf{a}_m^H(\theta) \mathbf{p}|, \quad (8)$$

where the  $M \times 1$  vector  $\mathbf{p}$  may be obtained as

$$\mathbf{p} = \mathbf{U}_a \Gamma \left( \frac{\mathbf{s}^H}{\|\mathbf{s}\|} \otimes \mathbf{U}_a^H \right) \mathbf{X}. \quad (9)$$

Here  $\Gamma = \text{diag}(\frac{1}{\|\mathbf{s}\|^2 \lambda_1 + \sigma^2}, \dots, \frac{1}{\|\mathbf{s}\|^2 \lambda_M + \sigma^2})$ . Recalling the form of  $\mathbf{a}_m(\theta)$  for a uniform linear array in (2), and again letting  $\phi = -\frac{\lambda_0 d}{2} \sin(\theta)$ , we see that  $T(\mathbf{X})$  may be expressed as  $T(\mathbf{X}) = M |P(\phi)|$ , where  $P(\phi)$  is the DTFT of the  $M \times 1$  vector  $\mathbf{p}$ .

#### 4. SPATIAL DECORRELATION VIA THE DFT

The previous section revealed that both the linear and quadratic terms require processing via the  $M \times M$  matrix  $\mathbf{U}_a^H$ . Unfortunately, spatially decorrelating the matched filter samples via  $\mathbf{U}_a^H$  is often computationally too intensive, especially for large arrays. In this section we propose a highly efficient method of spatially processing the observations. Note that if the distances between sensors are equal, as in a uniform linear array, then often  $\mathbf{R}_a$  is taken to be Toeplitz [10]. Based on this property, spatial decorrelation can be achieved asymptotically using the discrete Fourier transform<sup>1</sup> (DFT) [11]; that is, we may substitute the  $M \times M$  DFT matrix  $\mathbf{F}_M$  for  $\mathbf{U}_a^H$ . The basic idea is that as we increase the number of sensors, the matrix  $\mathbf{R}_a$  grows in dimension and asymptotically becomes a circulant matrix, which is diagonalized by the DFT matrix  $\mathbf{F}_M$ . Therefore, our approach will be to asymptotically approximate the optimal quadratic detector derived in the previous section by substituting the DFT matrix  $\mathbf{F}_M$  in place of the true spatial decorrelating matrix  $\mathbf{U}_a^H$ . This results in a detector employing only DFT's for spatial processing and simple matched filters in time. Since the cost of implementing a DFT is relatively small when FFT techniques are used, the computational expense of such a processor is comparable to that of existing traditional matched-field beamformers. Although the DFT approach is asymptotically optimal, we are also interested in how the decorrelating power of the DFT depends on the number of sensors  $M$  and

the covariance matrix  $\mathbf{R}_a$ . A good measure of the residual correlation still remaining is provided by the norm of the matrix containing the off-diagonal covariance elements of the transformed coefficients [11]. Defining  $\mathbf{D}_M$  as a diagonal matrix containing the same diagonal elements as the matrix  $\mathbf{F}_M \mathbf{R}_a \mathbf{F}_M^{-1}$ , a measure of residual correlation is given by the weak (Hilbert-Schmidt) norm of the difference matrix  $\mathbf{D}_M - \mathbf{F}_M \mathbf{R}_a \mathbf{F}_M^{-1}$ , defined by

$$\begin{aligned} \Gamma_M &= \|\mathbf{D}_M - \mathbf{F}_M \mathbf{R}_a \mathbf{F}_M^{-1}\|^2 \\ &= \frac{1}{M} \sum_{i=1}^M \sum_{j=1}^M (\|\mathbf{D}_M - \mathbf{F}_M \mathbf{R}_a \mathbf{F}_M^{-1}\|_{ij})^2. \end{aligned} \quad (10)$$

As for a DFT-based detector implementation, we can no longer use the eigenvalues  $\lambda_k$  of the matrix  $\mathbf{R}_a$  as we did in (5) and (8). We now have an approximate diagonal representation of  $\mathbf{R}_a$  given by  $\mathbf{R}_a \approx \sum_{k=1}^M \mu_k \mathbf{f}_k \mathbf{f}_k^H$  where  $\mathbf{f}_k$ ,  $k = 1, \dots, M$  are the columns of  $\mathbf{F}_M^{-1}$ . Note that  $\mu_k$  may be obtained simply as the DFT of the first row of  $\mathbf{R}_a$ . Therefore the DFT approximation to the quadratic term in (5) is given by

$$Q_{DFT}(\hat{\mathbf{Z}}) = \frac{1}{2\sigma^2} \sum_{k=1}^M \frac{\mu_k}{\|\mathbf{s}\|^2 \mu_k + \sigma^2} |\hat{z}_k|^2. \quad (11)$$

where  $\hat{\mathbf{Z}} = [\hat{z}_1, \dots, \hat{z}_M]^T = \mathbf{F}_M \mathbf{A}^H(\theta) (\mathbf{s}^H \otimes \mathbf{X})$  are the approximately spatially decorrelated, aligned matched filter samples. The DFT approximation to the linear part of the detector in (8) is obtained by making the same type of substitutions. Specifically,

$$T_{DFT}(\mathbf{X}) = |\mathbf{a}_m^H(\theta) \hat{\mathbf{p}}|, \quad (12)$$

where  $\hat{\mathbf{p}} = \mathbf{F}_M^{-1} \hat{\Gamma} \mathbf{F}_M (\mathbf{s}^H \otimes \mathbf{X})$ ,  $\hat{\Gamma} = \text{diag}(\frac{1}{\|\mathbf{s}\|^2 \mu_1 + \sigma^2}, \dots, \frac{1}{\|\mathbf{s}\|^2 \mu_M + \sigma^2})$ , and again the  $\mu_k$ 's are the DFT coefficients of the first row of  $\mathbf{R}_a$ .

The effect of the DFT substitution as an approximate decorrelator on the angle of arrival search is as follows. The approximately decorrelated, aligned matched filter samples are given by  $\hat{\mathbf{Z}} = \mathbf{F}_M \mathbf{A}^H(\theta) \mathbf{Y}$ . Analogous to (6), the  $l$ th element of this  $M \times 1$  snapshot may be expressed as

$$\hat{\mathbf{Z}}_l = \frac{1}{M} \sum_{k=1}^M e^{-j2\pi \frac{(k-1)(l-1)}{M}} y_k e^{-j2\pi(k-1)\phi} = Y(\phi + \frac{l-1}{M}), \quad (13)$$

where  $Y(\phi)$  is the DTFT of the matched filter samples  $\{y_k\}_{k=1}^M$ . That is,  $\hat{\mathbf{Z}} = [Y(\phi), Y(\phi + \frac{1}{M}), \dots, Y(\phi + \frac{M-1}{M})]^T$ . Therefore, the approximately decorrelated, aligned matched filter samples may be obtained efficiently as the DTFT of the matched filter samples with appropriate frequency shifts. In practice we would sample  $\phi = \frac{p}{M}$ ,  $p = 0, \dots, M-1$  to obtain a DFT, resulting in  $\hat{\mathbf{Z}}_l = [Y(p), Y(p+1), \dots, Y(p+M-1)]^T$ , where  $Y(p)$  now denotes the  $p$ th DFT sample of the matched filter samples. This reveals that the search over angle (now represented through the discrete variable  $p$ ) translates into simply circularly shifting the DFT coefficients of the matched filter samples. Note that multiplication of the magnitude-squared circularly shifted DFT coefficients with the weights  $\frac{\mu_k}{\|\mathbf{s}\|^2 \mu_k + \sigma^2}$  is simply the circular convolution of these two vectors. Hence fast convolution algorithms can be used to form the circular convolution, and choosing the maximum value of this convolution will result in the GLRT which incorporates the

<sup>1</sup>Note that the use of the DFT here is only for spatial decorrelation and is not related to angle-of-arrival searches.

unknown angle of arrival. As for dealing with the angle-of-arrival search with the linear term, we note that the inner product  $\mathbf{a}^H(\theta)\hat{\mathbf{p}}$  represents a particular DFT sample of the vector  $\hat{\mathbf{p}}$ . Therefore, from (12) a GLRT would require us to choose the largest element of the vector  $|\mathbf{F}_M \hat{\mathbf{p}}| = |\hat{\mathbf{F}}_M(\mathbf{s}^H \otimes \mathbf{X})| = |\hat{\mathbf{F}}\hat{\mathbf{Z}}|$ , where  $\hat{\mathbf{Z}} = \mathbf{F}_M(\mathbf{s}^H \otimes \mathbf{X}) = \{\hat{z}_k\}$  denotes the DFT of the matched-filter samples. The DFT-based implementation of the linear-quadratic detector is illustrated in Figure 1.

It is of interest to examine the form of the GLRT for the conventional matched-field beamformer which does not account for perturbations. In this case, the optimal test statistic is given by [6]

$$L_{MF}(\mathbf{X}) = |(\mathbf{a}_m^H(\theta) \otimes \mathbf{s}^H)\mathbf{X}|. \quad (14)$$

Again letting  $\mathbf{Y} = \mathbf{s}^H \otimes \mathbf{X}$  denote the matched filter samples and letting  $\phi = -\frac{f_0 d}{c} \sin(\theta)$ , we have  $L_{MF} = |Y(\phi)|$  where  $Y(\phi)$  is the DTFT of the matched filter samples. Hence a GLRT would simply require choosing the maximum value of magnitude of the DTFT coefficients. We may sample to obtain a DFT, and hence the GLRT may be realized by simply choosing the maximum DFT coefficient magnitude of the matched filter samples. The detector is illustrated in Figure 2. Comparing with Figure 1, note the only additional processing to deal with array miscalibration over conventional matched-field processing which ignores array miscalibration is the spatial smoothing provided by the circular convolution.

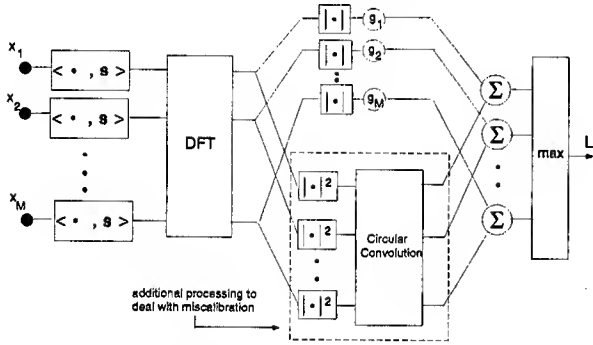


Figure 1: Implementation of the linear-quadratic detector via the DFT approximation; Here  $g_k = \frac{1}{\|\mathbf{s}\|^2 \mu_k + \sigma^2}$ ,  $k = 1, \dots, M$ , and the circular convolution block denotes the circular convolution of the input with the sequence  $\{\frac{1}{2\sigma^2} \frac{\mu_k}{\|\mathbf{s}\|^2 \mu_k + \sigma^2}\}$

## 5. SIMULATION

For simulation, we assume  $M = 64$  sensors in a ULA with half-wavelength spacing with the SNR of each sensor observation set to  $-40dB$ . We generated a radar waveform return using Matlab code modified from the Mountaintop Matlab toolbox<sup>2</sup>. The radar waveform consists of a burst of  $N_p = 16$  identical pulses with a pulse bandwidth of 500 kHz, pulse width of 100  $\mu s$ , PRI of 1.6 ms, and transmit frequency of 435 MHz. Reception was such that  $N = 6448$  samples were collected at each sensor. Array miscalibration was accounted for by assuming a symmetric Toeplitz covariance matrix for  $\mathbf{R}_a$  with first row equal to [6]:

$$\sigma_a^2 [1 \ \alpha \ \alpha^2 \ \dots \ \alpha^{M-1}], \quad (15)$$

<sup>2</sup>This toolbox is available at <ftp://ftp.ee.gatech.edu/pub/users/yaron/>

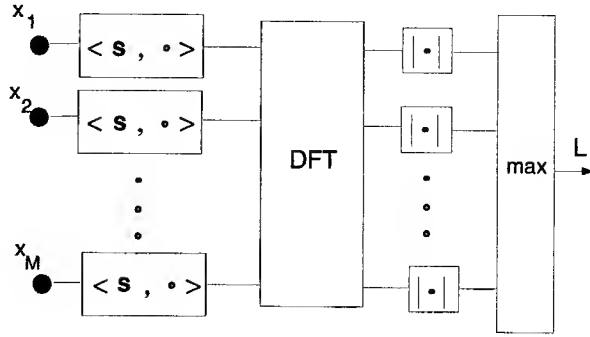


Figure 2: Implementation of the conventional matched-field processor incorporating a GLRT for angle search. This detector is optimal in the absence of array perturbations.

where  $\sigma_a^2$  is the variance of the array errors. Figure 3 shows the residual correlation  $\Gamma_M$  defined in (10) as a function of  $\alpha$ . Note that over a large range of  $\alpha$ , the residual correlation is quite low; the worst case occurs for  $\alpha = 0.95$ . A value of  $\sigma_a^2 = 0.5$  was used in the simulations. Figure 4 compares the receiver operating characteristic (ROC) for optimal combining, DFT combining, and matched filtering (which ignores array miscalibration), all for  $\alpha = 0.7$ . Note the significant gain in performance offered by DFT combining over matched filtering; the DFT approach performs virtually the same as optimal combining. Figure 5 illustrates the same information, only for  $\alpha = 0.95$ . Even in this worst-case scenario for  $\alpha$ , DFT combining significantly outperforms matched filtering, and there is only a slight loss in performance over optimal.

## 6. CONCLUSION

Fusing data collected from an array of sensors can considerably enhance signal detection, provided the data is processed in the proper fashion. Motivated by the fact that extensive previous studies have shown that random array perturbations can cause significant degradation in the performance of traditional matched-field beamformers which ignore random perturbations, we propose a quadratic array processor which fully incorporates the statistical nature of the perturbations. Implementation of the optimal detector requires two main steps: decorrelation in space followed by filtering in time. While deriving the form of the optimal quadratic processor is relatively straightforward, our main contribution is proposing an efficient method of implementing it. Recognizing that spatially decorrelating each snapshot is computationally quite expensive, we showed that spatial decorrelation can be approximately achieved in a general-purpose fashion with a discrete Fourier transform. We also illustrated how efficient frequency-domain techniques for angle-of-arrival searches can be easily incorporated into our proposed detector. It turns out that the suggested spatial DFT processing serves two purposes at once: efficient spatial decorrelation which deals with the random perturbations, and circular shifting of the DFT coefficients allows for a search over the unknown angle of arrival. Simulation results reveal that even in worst-case scenarios, the DFT approximation to the optimal quadratic detector not only significantly outperforms conventional matched filtering techniques, but provides near-optimal performance over a wide range of correlation in the perturbations. Hence our proposed implementation has a cost comparable to that of existing matched-

field beamformers while providing the performance benefits of the much more complicated quadratic processors.

## 7. REFERENCES

- [1] A. Swindlehurst, T. Kailath, "A Performance Analysis of Subspace-Based Methods in the Presence of Model Errors, Part I: The MUSIC Algorithm," *IEEE Transactions on Signal Processing*, vol. 40, no. 7, July 1992.
- [2] D. Morgan, "Coherence effects on the detection performance of quadratic array processors, with applications to large-array matched-field beamforming," *Journal of the Acoustical Society of America*, vol. 87, no. 2, pp. 737-747, February 1990.
- [3] S. Pillai, *Array Signal Processing*. New York: Springer-Verlag, 1989.
- [4] H. Krim, M. Viberg, "Two Decades of Array Signal Processing Research," *IEEE Signal Processing Magazine*, July 1996.
- [5] Y. Rockah, H. Messer, P.M. Schultheiss, "Localization Performance of Arrays Subject to Phase Errors," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 24, no. 4, pp. 402-410, July 1988.
- [6] S. Ricks, A. Swindlehurst, "Detection Performance Degradation due to Miscalibrated Arrays in Airborne Radar," *32nd Asilomar Conference on Signals, Systems, and Computers*, vol. 2, pp. 1532-1536, 1998.
- [7] B. Friedlander, "A Sensitivity Analysis of the MUSIC Algorithm," *IEEE Transactions on Acoust., Speech, and Signal Processing*, vol. 38, no. 10, October 1990.
- [8] M. Viberg, A. Swindlehurst, "A Bayesian Approach to Auto-Calibration for Parametric Array Signal Processing," *IEEE Transactions on Signal Processing*, vol. 42, no. 12, December 1994.
- [9] H. V. Poor, *An Introduction to Signal Detection and Estimation*. New York: Springer-Verlag, 1988.
- [10] A. Paulraj and T. Kailath, "Direction of arrival estimation by eigenstructure methods with imperfect spatial coherence of wave fronts," *Journal of the Acoustical Society of America*, vol. 83, pp. 1034-1040, March 1988.
- [11] J. Pearl, "On coding and filtering stationary signals by discrete Fourier transform," *IEEE Transactions on Information Theory*, IT-19, pp. 229-232, March 1973.

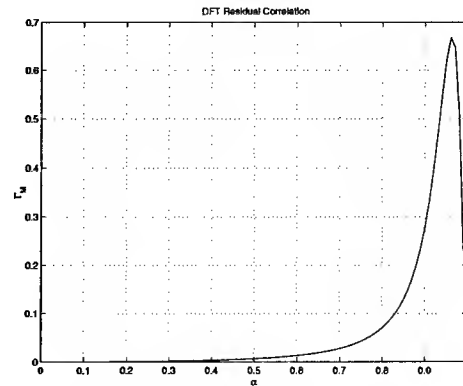


Figure 3: DFT residual correlation as a function of  $\alpha$ . Note that the residual correlation is small over a wide range of spatial correlation; the worst case occurs at  $\alpha = 0.95$ .

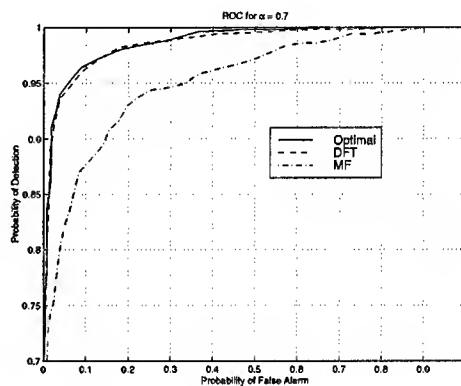


Figure 4: ROC for Optimal Combining, DFT Combining, and Matched Filter;  $\alpha = 0.7$ . Note not only the significant performance gain, but the near-optimal performance provided by DFT combining.

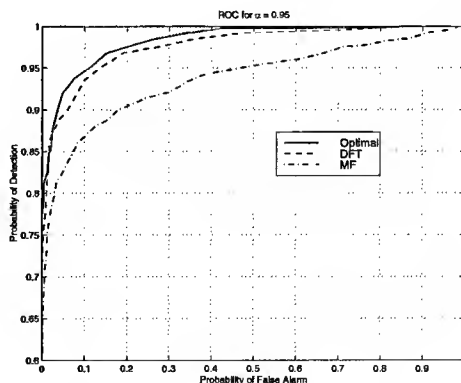


Figure 5: ROC for Optimal Combining, DFT Combining, and Matched Filter;  $\alpha = 0.95$ . Even in this worst-case scenario, DFT combining significantly outperforms matched filtering.

# A NEURAL NETWORK APPROACH FOR DOA ESTIMATION AND TRACKING

L. Badidi and L. Radouane

LESSI, Département de physique  
Faculté des Sciences B. P. 1796  
Fes-Atlas-30000  
Morocco  
E-mail: [lbadidi@hotmail.com](mailto:lbadidi@hotmail.com)

**Abstract**— Many signal subspace based approaches have been proposed for determining the fixed direction of arrival (DOA) of plane waves impinging on an array of sensors. However, computational burden of subspace based algorithms makes them unsuitable for real time processing of nonstationary signal parameters. In this work, we present an iterative procedure for DOA estimation and tracking. The complete procedure consists in, first, extracting the noise or signal subspace, by training the MCA or PCA algorithms, respectively. These algorithms contain only relatively simple operations and have self-organizing properties. Then, using Newton algorithm, we get the estimated DOA. The performance on simulated data representing both constant and time-varying signals of the approach is presented.

## I. Introduction

Subspace-based methods for estimating the frequencies of sinusoids or the DOA of signals impinging on an array of sensors have drawn considerable interest over recent years. State-space method [1], ESPRIT [2], MUSIC [3], and Min-Norm [4] are examples of these techniques. Based on the eigendecomposition of covariance matrix of the array output, they offer high resolution and give accurate estimate [5]. A key limitation of these techniques is the computational burden to process a new sample (snapshot), so they are unsuitable for real time applications.

Some attempts have been made to reduce the computational burden of these methods. Stewart [6] has introduced the URV decomposition developed by Liu et al. [2]. Eriksson et al. [7] extended the method called subspace estimation without eigendecomposition (SWEDE) studied in [8], the method estimates the DOA by linear operation on the data. An alternative procedure is to use some adaptive algorithm for on-line estimation of the desired subset eigendata [9]- [16].

In this work, we focus on the MUSIC method of Schmidt [3]; this method involves solving a one-dimensional minimization problem and finding subspace (noise or signal subspace).

To deal with the computational complexity of the subspace based method, we present an iterative procedure to update the DOA. The complete procedure consists of two steps: one performs the extraction of the noise (or signal) subspace using the well-known Oja (or anti-Hebbian) algorithm [13], respectively. Then, the second performs the one-dimensional minimization using the Newton algorithm [7].

The rest of this paper is organized as follows. In Section II, we formulate the problem. The PCA, MCA learning and the Newton algorithms are given in Section III. Then, in Section IV, we present simulation results. Finally, Section V summarizes our conclusions.

## II. PROBLEM FORMULATION AND BASIC ASSUMPTIONS

Consider a linear array of  $N$  sensors. The array output is commonly modeled as follows [1]

$$\mathbf{x}(t) = \mathbf{D}(\theta)\mathbf{s}(t) + \mathbf{v}(t) \quad (1)$$

where  $\mathbf{D}(\theta) = [\mathbf{d}(\theta_1), \dots, \mathbf{d}(\theta_p)]$  is a  $N \times p$  matrix whose columns are the direction vectors with parameter vector  $\theta$  denoting the angles of arrival of the  $p$  signals.  $\mathbf{s}(t)$  is a  $p \times 1$  vector which denotes the complex envelopes of the narrow-band signals. The elements of  $\mathbf{s}(t)$  are assumed to be independent Gaussian distributed random variables with zero mean.  $\mathbf{v}(t)$  is a  $N \times 1$  vector representing the receiver noise of the  $N$  sensors. It is assumed to be complex, zero mean, Gaussian white process, and independent of the signals.

Form (1), the covariance matrix of  $\mathbf{x}(t)$  is given by

$$\mathbf{R} = E[\mathbf{x}(t)\mathbf{x}^H(t)] = \mathbf{D}(\theta)\mathbf{R}_s\mathbf{D}^H(\theta) + \sigma^2\mathbf{I} \quad (2)$$

where  $\sigma^2$  is an unknown constant representing the noise power,  $\mathbf{I}$  is an identity matrix of appropriate dimension, and  $(\cdot)^H$  denotes conjugate transpose.

$\mathbf{R}_s = E[\mathbf{s}(t)\mathbf{s}^H(t)]$  is the signal covariance matrix. Under the assumption of incoherent signals, the rank of  $\mathbf{R}_s$  is

$p'=p$ , The eigendecomposition of the positive definite Hermitian matrix  $\mathbf{R}$  is given by

$$\mathbf{R} = \sum_{i=1}^p \lambda_i \mathbf{e}_i \mathbf{e}_i^H + \sigma^2 \sum_{i=p+1}^N \mathbf{e}_i \mathbf{e}_i^H \quad (3)$$

where  $\lambda_i$  is the eigenvalue corresponding to the eigenvector  $\mathbf{e}_i$ , stored in decreasing order, for all  $i, \dots, N$ . Let  $\mathbf{W}_J = \{\mathbf{e}_1, \dots, \mathbf{e}_p\}$  and

$\mathbf{W}_v = \{\mathbf{e}_{p+1}, \dots, \mathbf{e}_N\}$  are the signal and noise subspaces,

respectively. From (2), (3), we have  $\mathbf{R}\mathbf{W}_v = \sigma^2 \mathbf{W}_v$ , then

$$\mathbf{D}^H(\theta) \mathbf{W}_v = 0, \quad \text{or} \quad \text{equivalently}$$

$\mathbf{d}^H(\theta_i) \mathbf{W}_v = 0$ , for  $i = 1, \dots, p$ . Hence, consistent estimate of the DOA's can be determined as the minimizing arguments of the cost function

$$f(\theta) = \mathbf{d}^H(\theta) \mathbf{\Pi} \mathbf{d}(\theta) \quad (4)$$

where  $\mathbf{\Pi} = \mathbf{W}_v \mathbf{W}_v^H = \mathbf{I} - \mathbf{W}_s \mathbf{W}_s^H$ .

In practical applications, the exact ensemble covariance matrix is not known. A solution to this problem consists in estimating the covariance matrix from a finite number

of snapshots  $\{\mathbf{x}(1), \mathbf{x}(2), \dots, \mathbf{x}(T)\}$

$$\hat{\mathbf{R}} = \frac{1}{T} \sum_{k=1}^T \mathbf{x}(k) \mathbf{x}^H(k) \quad (5)$$

Hence, consistent estimates of the DOA's can be obtained from the following three steps:

Computation of  $\hat{\mathbf{R}}$

Eigendecomposition of  $\hat{\mathbf{R}}$

Minimization of  $f(\theta)$  (equation 4).

The computational burden needed by the eigendecomposition of the matrix  $\hat{\mathbf{R}}$  and the minimization of (4) makes the MUSIC algorithm unsuitable in the nonstationary case. To deal with this problem, in next section, we present a PCA or MCA for extracting the signal subspace or noise subspace, respectively. Then, to get the estimated DOA's, we use the Newton algorithm for minimizing the cost function (4).

### III. Derivation of the Algorithms

The eigenvectors corresponding to the largest and smallest eigenvalues of the autocorrelation matrix of the input signals are referred as the principal components and minor components, respectively [9-14]. Adaptively extracting the minor and principal components is a primary requirement in many fields of signal processing, including the eigen-based bearing estimation (MUSIC and Min-Norm...). The main purpose of this subsection is to present the Oja or the anti-Hebbian algorithm for

extracting the principal components or the minor components, respectively, and the Newton algorithm for minimizing the cost function (4).

#### A. PCA & MCA Algorithms

The set of  $M$  ( $M = N - p$ ) minimum eigenvectors (minor components) of  $\mathbf{R}$  can be written as the solution of the following constrained minimization problem [9]-[14]:

$$\min_{\mathbf{e}_k} \mathbf{e}_k^H \bar{\mathbf{R}} \mathbf{e}_k \quad \text{subject to} \quad \mathbf{e}_k^H \mathbf{e}_l = \delta_{kl} \quad (6)$$

$$\forall k, l \in \{1, \dots, M\}$$

where  $\delta_{kl}$  is the Kronecker delta function. Equation (6) is a complex-value constrained quadratic programming problem. To convert it into a real-value constrained quadratic programming formulation, the complex vectors  $\mathbf{e}_k$ ,  $k = 1, \dots, M$  and  $\mathbf{R}$  should first be decomposed into their real and imaginary constituents as follows [12]:

$$\mathbf{e}_k = \mathbf{e}_{kr} + j \mathbf{e}_{ki} \quad \text{and} \quad \mathbf{R} = \mathbf{R}_r + j \mathbf{R}_i \quad (7)$$

then the equation becomes

$$[\mathbf{R}_r + j \mathbf{R}_i][\mathbf{e}_{kr} + j \mathbf{e}_{ki}] = \lambda_k [\mathbf{e}_{kr} + j \mathbf{e}_{ki}] \quad (8)$$

or equivalently,

$$\begin{aligned} \mathbf{R}_r \mathbf{e}_{kr} - \mathbf{R}_i \mathbf{e}_{ki} + j [\mathbf{R}_r \mathbf{e}_{ki} + \mathbf{R}_i \mathbf{e}_{kr}] \\ = \lambda_k (\mathbf{e}_{kr} - \mathbf{e}_{ki} + j [\mathbf{e}_{ki} + \mathbf{e}_{kr}]) \end{aligned} \quad (9)$$

Moreover, by combining terms, we get

$$\mathbf{R}_c \mathbf{w}_k = \lambda_k \mathbf{w}_k \quad (10)$$

with

$$\mathbf{R}_c = \begin{pmatrix} \mathbf{R}_r & -\mathbf{R}_i \\ \mathbf{R}_i & \mathbf{R}_r \end{pmatrix} \quad \text{and} \quad \mathbf{w}_k = \begin{pmatrix} \mathbf{e}_{kr} \\ \mathbf{e}_{ki} \end{pmatrix} \quad (11)$$

where, we have used

$$\begin{aligned} \mathbf{R}_c &= E[\mathbf{X}_c(t) \mathbf{X}_c^T(t)], \\ \mathbf{R}_r &= E \left[ \mathbf{x}_r(t) \mathbf{x}_r^T(t) \right] + E \left[ \mathbf{x}_i(t) \mathbf{x}_i^T(t) \right], \\ \mathbf{R}_i &= E \left[ \mathbf{x}_i(t) \mathbf{x}_r^T(t) \right] - E \left[ \mathbf{x}_r(t) \mathbf{x}_i^T(t) \right], \end{aligned}$$

$$\text{and } \mathbf{X}_c = \begin{pmatrix} \mathbf{x}_r & -\mathbf{x}_i \\ \mathbf{x}_i & \mathbf{x}_r \end{pmatrix}.$$

$\mathbf{R}_c$  is a  $2N \times 2N$  symmetric, positive-definite matrix, and  $\mathbf{w}_k$  is a  $2N \times 1$  column vector. Therefore, the complex-value constrained minimization problem (6) becomes

$$\min_{\mathbf{w}_k} \mathbf{w}_k^T \mathbf{R}_c \mathbf{w}_k \text{ subject to } \mathbf{w}_k^T \mathbf{w}_l = \delta_{kl} \quad (12)$$

$$\forall k, l \in \{1, \dots, M\}$$

The solution of the minimization problem (12) is given by [13]:

$$\mathbf{W}_{k+1} = \mathbf{W}_k - \alpha_k \left[ \mathbf{I} - \mathbf{W}_k \mathbf{W}_k^T \right] \mathbf{X}_c(k) \mathbf{X}_c^T(k) \mathbf{W}_k \quad (13)$$

This algorithm is known as the anti-Hebbian algorithm [12]. The algorithm (13) can be used to extract the  $M$  ( $M = p$ ) principal components, simply by changing the sign of the parameter  $\alpha_k$  as

$$\mathbf{W}_{k+1} = \mathbf{W}_k + \alpha_k \left[ \mathbf{I} - \mathbf{W}_k \mathbf{W}_k^T \right] \mathbf{X}_c(k) \mathbf{X}_c^T(k) \mathbf{W}_k \quad (14)$$

This algorithm can be also obtained by minimizing the mean-square representation error [14]. The minor and principal subspace algorithms (13-14) can be modified to a more general form as [12], [15], and [16]

$$\mathbf{W}_{k+1} = \mathbf{W}_k \mp \alpha_k \left[ \mathbf{I} - \beta_k \mathbf{W}_k \mathbf{W}_k^T \right] \mathbf{X}_c(k) \mathbf{X}_c^T(k) \mathbf{W}_k \quad (15)$$

where  $\beta_k$  is a positive parameter. This algorithm is called the weighted subspace algorithm [12].

### B. Newton Algorithm [7]

In the stationary case, the minimum of  $f(\theta)$  can be located as the peaks of the spectrum (4). However, in the nonstationary case, to track the minimum of the cost function  $f(\theta)$ , the following approximate Newton algorithm [7], can be used

$$\hat{\theta}_l = \bar{\theta}_l - \frac{\text{Re}[\partial d_l^H \Pi d_l]}{\partial d_l^H \Pi \partial d_l} \bigg|_{\theta_j = \bar{\theta}_j} \quad (16)$$

$$\partial d_l(m) = j\pi(m-1) \cos(\theta_l) e^{j\pi \sin(\theta_l)},$$

$$m = 1, \dots, N$$

where  $\text{Re}(\cdot)$  is the real part, and  $\left\{ \bar{\theta}_l \right\}_{l=1}^P$  are the most

recent minimum estimates available. This Newton step might be reiterated a few times at each minimum updating step, starting from the estimates provided by the previous update. This algorithm can be easily derived, using the Taylor expansion, by taking  $f'(\theta) = 0$  at the optimum, for further information see [7].

The proposed procedure in this paper, can be summarized as follows:

- 1) Initialize  $\mathbf{W}_0$  as a  $2N \times M$  random matrix whose columns are orthogonal and with unit norm.

- 2) Update  $\mathbf{W}$  as in equation (15).

- 3) Update  $\theta_i$ ,  $i = 1, \dots, p$  as in (16).

## IV. Computer simulations

In this section, we present some simulation results illustrating the properties of the proposed approach. In all examples, we use a uniform linear array with 12 elements spaced  $\lambda/2$  apart, where  $\lambda$  denotes the wavelength of the sources signals.

The steering vector  $\mathbf{d}(\theta)$  is then given by

$$\mathbf{d}(\theta) = [e^{j\pi \sin(\theta)}, \dots, e^{j\pi(n-1) \sin(\theta)}]^T, \quad j = \sqrt{-1}.$$

The noise is modeled as complex Gaussian with zero mean and variance  $\sigma^2$  for all sensors. The signal to noise

ratio is defined as  $10 \log(\frac{a_i^2}{\sigma^2})$ . We consider the estimation of the DOA in the following scenarios: ones involving close sources located at:

$$10^\circ + 5^\circ \sin(2\pi t / 360) \quad (17)$$

$$20^\circ + 5^\circ \sin(2\pi t / 240), \quad t = 1, 2, \dots, 700$$

second one involving well-separated sources located at

$$-5^\circ + 10^\circ \sin(2\pi t / 360) \quad (18)$$

$$40^\circ + 5^\circ \sin(2\pi t / 240), \quad t = 1, 2, \dots, 700$$

and finally, the third is concerned with instantaneously changing sources. We have assumed that there were two sources located at  $14^\circ$  and  $17^\circ$ , each with SNR=26 dB, and that the signals alternatively appear and disappear.

This example is adopted from [12], because it corresponds to a sampling rate of 1 data point per  $.11^\circ$ ,  $.08^\circ$  or  $.05^\circ$  change in angle, and typical radar applications produce 1 point per  $10^{(10-5)^\circ}$ , so this example is much more demanding.

We have simulated the above iterative procedure using the learning rule (15), with  $\beta = 0.1$  and  $\alpha = 0.01$ , to extract the noise subspace. The initial weight matrix  $\mathbf{W}_0$  is chosen to satisfy  $\mathbf{W}_0^T \mathbf{W}_0 = \mathbf{I}$ , then the minima of the cost function (4) are computed using the algorithm (16).

Fig. 1 gives the result for the well-separated sources, Fig. 2 gives the result for the closed sources, and Fig. 3 gives the results for the instantaneous changing sources.

Another example concerns two fixed signal sources located at  $24^\circ$  and  $29^\circ$ , the source power was 26 and 23 dB above the background noise. The gain parameter  $\alpha_k = 0.02$  was constant during the first 50 iterations and then decreased slowly. The result is given in Figs. 4-5.



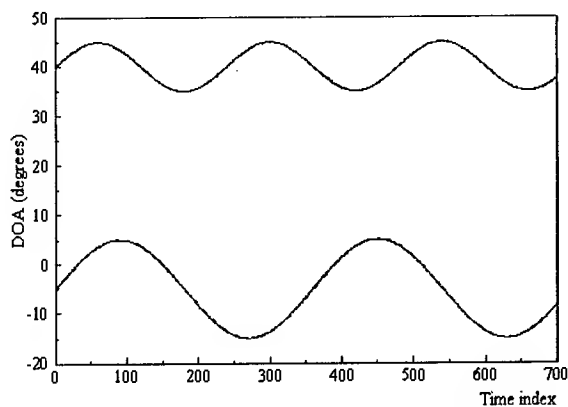


Fig. 1 Estimated time-varying DOA's for well-separated sources.

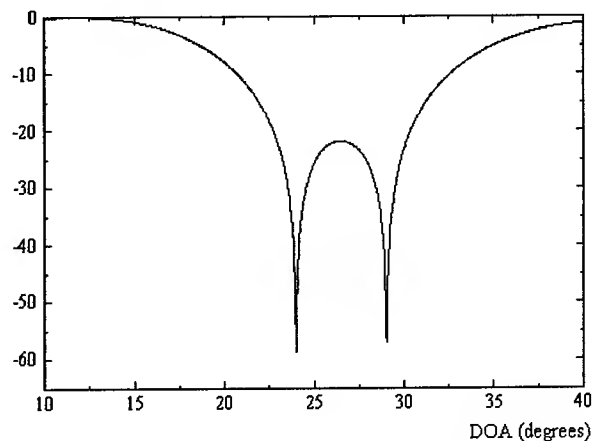


Fig. 4 The estimated spectrum versus DOA (degrees).

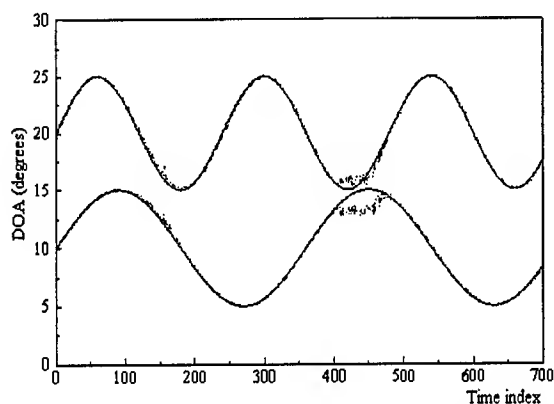


Fig. 2 Estimated time-varying DOA's for closed sources.

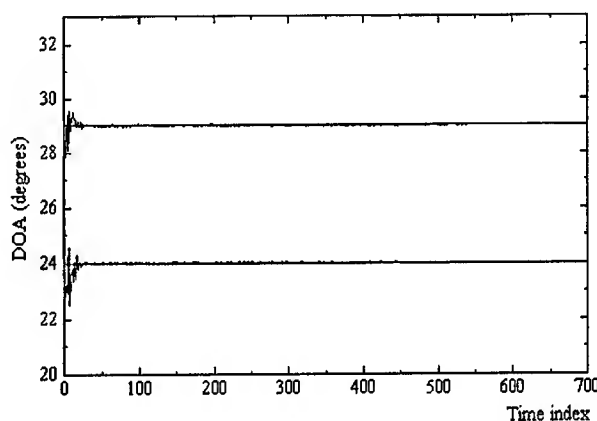


Fig. 5 Estimated DOA for fixed sources versus  $\theta$  (degree).

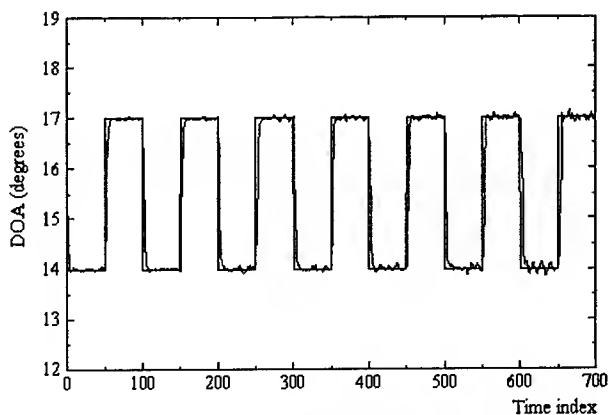


Fig. 3 Estimated time-varying DOA's for instantaneous changing sources.

Simulation results show that the proposed approach can be successfully used for real-time tracking of time varying signals.

## V. Conclusion

This paper presented an iterative procedure to DOA estimation and tracking. The main purpose of the paper is to deal with the computational complexity of the subspace based methods. Indeed, to alleviate the eigendecomposition of the covariance matrix the anti Hebbian algorithm is trained to extract the noise subspace. Then, the Newton algorithm is used to perform the one dimensional minimization problem. The performance of the approach is supported by numerical experiments.

## References

- [1] R. J. Vaccaro and Y. Ding, "A new state-space approach for direction finding," *IEEE Trans. Signal Processing*, vol. 42, no. 11, pp. 3234-3237, Nov. 1994.
- [2] K. J. R. Liu, D. P. O'Leary, G. W. Stewart, and Y.-J. J. Wu, "URV ESPRIT for tracking time-varying signals," *IEEE Trans. Signal Processing*, vol. 42, no. 12, pp. 3441-3448, Dec. 1994.
- [3] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propaga.*, vol. 34, no. 3, pp. 276-280, Mar. 1986.
- [4] R. Kumaresan and D. W. Tufts, "Estimating the angles of arrival of multiple source plane waves," *IEEE Trans. Aerosp. Elect. Syst.*, vol. AES-19, pp. 134-149, January, 1983.
- [5] M. Viberg and B. Ottersten, "Sensor array processing based on subspace fitting," *IEEE Trans. Signal Processing*, vol. 39, no. 5, pp. 1110-1120, May 1989.
- [6] G. W. Stewart, "An updating algorithm for subspace tracking," *IEEE Trans. Signal Processing*, vol. 40, pp. 1535-1541, 1992.
- [7] A. Eriksson, P. Stoica, and T. Söderström, "On-line subspace algorithms for tracking moving sources," *IEEE Trans. Signal Processing*, vol. 42, no. 9, pp. 2319-2330, Sept. 1994.
- [8] C. C. Yeh, "Simple computation of projection matrix for bearing estimation," *Proc. Inst. Elec. Eng.*, Part F, vol. 134, pp. 146-150, Apr. 1987.
- [9] J.-F. Yang, and M. Kaveh, "Adaptive eigensubspace algorithms for direction or frequency estimation and tracking," *IEEE Trans., Acoustics, Speech, Signal Processing*, vol. ASSP-36, no. 2, pp. 241-251, Feb. 1988.
- [10] J. Karhunen, "Recursive estimation of eigenvectors of correlation type matrices for signal processing application," Ph. D Dissertation, Helsinki Univ., Finland. 1984.
- [11] E. Oja and J. Karhunen, "On stochastic approximation of the eigenvectors and eigenvalues of the expectations of a random matrix," *J. of Math. Stat. Analysis and Applications*, vol. 106, pp. 69-84, 1985.
- [12] F.-L. Luo, and R. Unbehauen, *Applied neural networks for signal processing*, Cambridge: University Press, 1997.
- [13] E. Oja, "Principal components, minor components, and linear neural networks," *Neural Networks*, vol. 5 pp. 927-935, 1992.
- [14] J. Karhunen and J. Joutsensalo, "Representation and separation of signals using nonlinear PCA type learning," *Neural Networks*, vol. 7 pp. 113-127, 1994.
- [15] E. Oja, H. Ogawa, and J. Wangviattana, "Principal component analysis by homogeneous neural networks, part I: the weight subspace criterion," *IEICE Trans.*, E-75D, pp. 366-375, 1992.
- [16] E. Oja, H. Ogawa, and J. Wangviattana, "Principal component analysis by homogeneous neural networks, part II: analysis and extensions of the learning," *IEICE Trans.*, E-75D, pp. 376-382, 1992.

# PARTIALLY ADAPTIVE ARRAY ALGORITHM COMBINED WITH CFAR TECHNIQUE IN TRANSFORM DOMAIN

*Sung-Hoon Moon, Dong-Hyun Yun, and Dong-Seog Han*

School of Electronic and Electrical Engineering, Kyungpook National University  
Taegu 702-701, Korea

## ABSTRACT

A frequency domain partially adaptive algorithm, called a censoring adaptive array (CAA) algorithm, is proposed to reduce the computational complexity of a frequency domain adaptive array. The CAA algorithm uses a cell averaging constant false alarm rate (CA-CFAR) processor to adapt only those weights that correspond to the frequency bins expected to contain interferences. The false alarm rate also adapts according to the environment. Furthermore, a censoring spatial smoothing is proposed to combine the CAA algorithm with a spatial smoothing technique. The performances of the proposed algorithms are compared with conventional algorithms using computer simulation.

## 1. INTRODUCTION

Adaptive array systems are designed to obtain only a desired signal in interference conditions by producing a null pattern or reducing the sidelobe level to the incident angles of the interference. Normally, conventional adaptive arrays have two critical problems. The first is that the required time for the convergence of the array output to a stable level is very long when the eigenvalues of the input correlation matrix are widely spread out [1]. Therefore, to increase the convergence speed, the adaptive array must remove the correlation between the tap-input signals of each tapped delay line (TDL) so that the eigenvalue spread is minimized. Chen and Fang [2] used a frequency domain least mean square (LMS) algorithm including a self-orthogonalizing property to remove the temporal correlation effectively. In contrast, An and Champagne [3] used a two-dimensional transform that can remove both temporal and spatial correlations. However, the computational complexity of the above algorithms is very high because of the computation involved in transforming the input signals into the frequency domain. Therefore, the computational complexity of frequency domain adaptive arrays needs to be reduced in order to make the systems more practical. The second problem

is a signal cancellation phenomenon caused by coherent interferences or smart jammers [4]. This signal cancellation phenomenon occurs whenever the interferences are correlated with the desired signal, thereby resulting in signal loss and severe signal distortion for narrow band and wide band signals, respectively. Spatial smoothing has been widely adopted to solve the signal cancellation phenomenon. However, the computational complexity of an array with spatial smoothing is much higher than that of an array without smoothing because the original array configuration is changed to many subarrays.

Accordingly, this paper proposes a new frequency domain partially adaptive algorithm, called a censoring adaptive array (CAA) algorithm, which can reduce the computational complexity of frequency domain adaptive algorithms while maintaining the performances of fully adaptive algorithms. When the CAA algorithm is combined with spatial smoothing, this can solve both the computational complexity problem in frequency domain adaptive algorithms and the signal cancellation phenomenon for coherent interferences.

The conventional frequency domain adaptive algorithm is described in section 2. Section 3 presents the CAA algorithm combined with spatial smoothing to remove coherent interferences. The simulation results are shown in section 4, and some concluding remarks are made in section 5.

## 2. CONVENTIONAL ALGORITHM

The convergence speed of a time domain adaptive array is very slow when the eigenvalues of the input correlation matrix are widely spread out. To increase the convergence speed, a frequency domain adaptive array has been proposed [2]. The frequency domain generalized sidelobe canceller (GSC) utilizing the frequency domain LMS algorithm is shown in Fig. 1. Unlike the Griffiths-Jim GSC, transform matrix  $\mathbf{D}$  is inserted after the blocking matrix in the auxiliary channel. The goal of the frequency domain GSC is to increase the con-

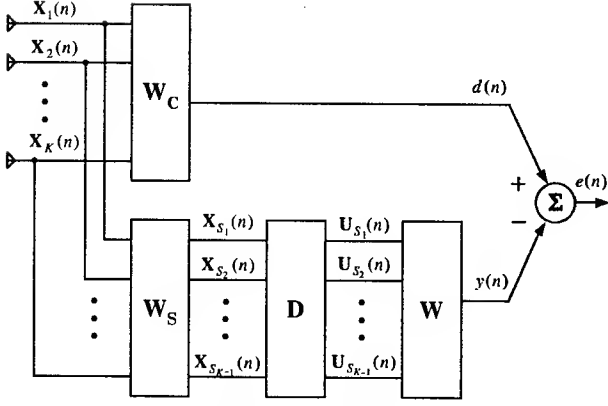


Figure 1: Block diagram of the frequency domain GSC.

vergence speed of the Griffiths-Jim GSC. To apply the frequency domain LMS algorithm, the discrete Fourier transform (DFT) transforms the input signals into the frequency domain and removes any correlation between the input signals. If it is assumed that GSC has  $K$  antenna elements and the length of TDL is  $L$ , then the array input signal vector,  $\mathbf{X}(n)$ , at the  $n$ th iteration time can be given by

$$\mathbf{X}(n) = [\mathbf{X}_1^T(n), \mathbf{X}_2^T(n), \dots, \mathbf{X}_K^T(n)]^T \quad (1)$$

where  $\mathbf{X}_i(n) = [x_i(n), x_i(n-1), \dots, x_i(n-L+1)]^T$  is the input signal vector of the  $i$ th antenna element, and  $x_i(j)$  is the input signal of the  $i$ th antenna element at the  $j$ th adaptation cycle. The superscript  $T$  denotes the transpose.  $d(n)$  and  $y(n)$  are the main and auxiliary channel output signals, respectively. If the signal blocking matrix consists of simple difference functions, the output vector of the signal blocking matrix,  $\mathbf{X}_S(n)$ , can be expressed as

$$\mathbf{X}_S(n) = [\mathbf{X}_{S1}^T(n), \mathbf{X}_{S2}^T(n), \dots, \mathbf{X}_{S(K-1)}^T(n)]^T \quad (2)$$

where  $\mathbf{X}_{Si}(n)$  is the input signal vector of the  $i$ th TDL. The output signal of the frequency domain GSC can be expressed as

$$e(n) = (\mathbf{W}_C^H - \mathbf{W}^H \mathbf{D} \mathbf{W}_S) \mathbf{X}(n). \quad (3)$$

The optimum weight vector of the frequency domain GSC is similar to that of the Griffith-Jim GSC given and can be expressed by

$$\mathbf{W}_{opt} = \mathbf{R}_{U_S}^{-1} \mathbf{P}_{U_S} \quad (4)$$

where  $\mathbf{U}_S(n) = [u_{1,1}(n), u_{2,1}(n), \dots, u_{L,K-1}(n)]^T$  is the transform domain vector of  $\mathbf{X}_S(n)$ , which can be expressed as

$$\mathbf{U}_S(n) = \mathbf{D} \mathbf{X}_S(n). \quad (5)$$

Then the auto-correlation matrix  $\mathbf{R}_{U_S}$  and the cross-correlation matrix  $\mathbf{P}_{U_S}$  are  $\mathbf{R}_{U_S} = E[\mathbf{U}_S(n) \mathbf{U}_S^H(n)]$ , and  $\mathbf{P}_{U_S} = E[\mathbf{U}_S(n) d^*(n)]$ , respectively. By using (5),  $\mathbf{R}_{U_S}$  and  $\mathbf{P}_{U_S}$  can be denoted as

$$\mathbf{R}_{U_S} = \mathbf{D} \mathbf{R}_{X_S} \mathbf{D}^H \quad (6)$$

and

$$\mathbf{P}_{U_S} = \mathbf{D} \mathbf{P}_{X_S}. \quad (7)$$

Then the mean square error (MSE) of the frequency domain GSC can be expressed as

$$\zeta = E[|e(n)|^2] = \sigma_d^2 - \mathbf{P}_{U_S}^H \mathbf{R}_{U_S}^{-1} \mathbf{P}_{U_S} \quad (8)$$

where  $\sigma_d^2$  is the variance of  $d(n)$ . By inserting (6) and (7) into (8), it can be easily shown that the MSE of the frequency domain GSC is the same as that of the Griffiths-Jim GSC. Hence, the MSE can be re-expressed as

$$\begin{aligned} \zeta &= \sigma_d^2 - (\mathbf{D} \mathbf{P}_{X_S})^H (\mathbf{D} \mathbf{R}_{X_S} \mathbf{D}^H)^{-1} (\mathbf{D} \mathbf{P}_{X_S}) \\ &= \sigma_d^2 - \mathbf{P}_{X_S}^H \mathbf{R}_{X_S}^{-1} \mathbf{P}_{X_S}. \end{aligned} \quad (9)$$

When the eigenvalues of the input correlation matrix are widely spread out, the convergence speed of the frequency domain adaptive array is much faster than that of the time domain adaptive array. However, it involves a high computational complexity when transforming the input signals into the frequency domain.

### 3. PROPOSED ALGORITHM

#### 3.1. Censoring Adaptive Algorithm

After the input signals are transformed into the frequency domain, the CAA algorithm uses a cell averaging constant false alarm rate (CA-CFAR) processor [5] to determine the frequency bins which contain components of the interference signals. A block diagram of the CA-CFAR processor is shown in Fig. 2. In the CA-CFAR processor, the threshold,  $Z_T$ , which is adjusted for each frequency bin, is obtained as follows:

$$Z_T = b \left( P_{FA}^{-\frac{1}{L-1}} - 1 \right) (L-1)^{-1} \quad (10)$$

where  $P_{FA}$ ,  $b$ , and  $L$  are the desired false alarm rate for the detection of frequency bins containing interfering signals, the sum of the neighboring data except for those frequency bins being tested, and the length of the TDL for each array element, respectively. After the CA-CFAR processor has tested all the frequency bins, the frequency domain adaptive algorithm only updates the weights connected to the frequency bins whose contents are greater than the threshold  $Z_T$ .

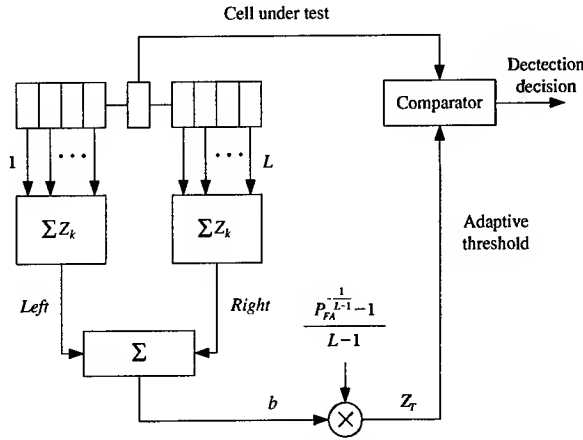


Figure 2: Block diagram of the CA-CFAR processor.

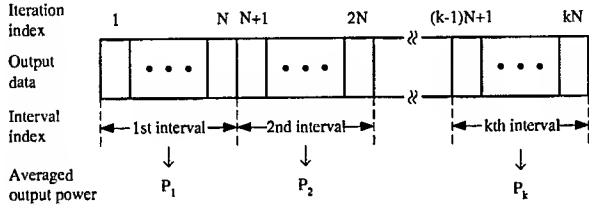


Figure 3: Operation for monitoring variations in output powers.

In order to benefit from the advantages of partial adaptive processing, the false alarm rate of the CAA algorithm must be changed adaptively according to environment variations. Therefore, the CAA algorithm monitors any variation in the output power, as shown in Fig. 3, so as to change the false alarm rate adaptively. The averaged output power  $M_k$  at the  $k$ th observation interval is obtained by  $M_k = \frac{1}{N} \sum_{n=1}^N O_{(k-1)N+n}$ , where  $O_i$  is the output power at the  $i$ th iteration. One observation interval consists of  $N$  iterations for operating the frequency domain LMS algorithm. The CAA algorithm then changes the false alarm rate of the CFAR processor adaptively using these averaged output powers such as  $P_{FA}(k) = P_{FA}(k-1) + \gamma S_k$ , where  $\gamma$  is the scaling constant,  $P_{FA}(k)$  is the false alarm rate at the  $k$ th observation interval, and  $S_k = M_k - M_{k-1}$ . As a result, the CAA algorithm is able to adaptively determine the optimum subspace for performing the frequency domain partial adaptation relative to the environment plus substantially reduce the computational complexity required for the weight adaptation.

Table 1: Computational complexities of GSCs for different algorithms. ( $K$ : number of antenna elements,  $L$ : length of TDL,  $L_i$ : number of updated weights in  $i$ th TDL, Transform method: FFT)

Algorithm	Complex multiplication/cycle
LMS	$2L(K-1)$
FLMS	$L(K-1) \log_2 L + 3.5L(K-1)$
CAA	$L(K-1) \log_2 L + 3.5 \sum_{i=1}^{K-1} L_i$

### 3.2. Mathematical Description

The Wiener-Hopf equation that denotes the optimum weight vector,  $\mathbf{W}_{opt}$ , of the full rank GSC is given by

$$\mathbf{R}_{\mathbf{U}_S} \mathbf{W}_{opt} = \mathbf{P}_{\mathbf{U}_S} \quad (11)$$

where  $\mathbf{R}_{\mathbf{U}_S}$  and  $\mathbf{P}_{\mathbf{U}_S}$  are the auto-correlation matrix of the transformed input data,  $\mathbf{U}_S(n)$ , and the cross-correlation matrix between  $\mathbf{U}_S(n)$  and the main channel output signal,  $d(n)$ , respectively.

Based on the self-orthogonality property of the frequency domain LMS algorithm, all the diagonal terms of  $\mathbf{R}_{\mathbf{U}_S}$  are equal [2]. Accordingly, the larger the value of  $(\mathbf{P}_{\mathbf{U}_S})_{i,j}$  which is the cross-correlation between  $d(n)$  and the frequency domain signal at the  $j$ th frequency bin in the  $i$ th TDL, the more it affects the MSE performance of the GSC [6]. Hence, the subspace for partial adaptation is composed of frequency bins that have a large value of  $(\mathbf{P}_{\mathbf{U}_S})_{i,j}$ . In other words, the signal subspace composed of the signals in the frequency bins that have a high correlation with the main channel signal can be the optimum subspace for minimizing the MSE. The proposed GSC adopting the CAA algorithm is shown in Fig. 4. The proposed GSC uses the CFAR processor to select the subspace composed of the signals in the frequency bins that have a high correlation with

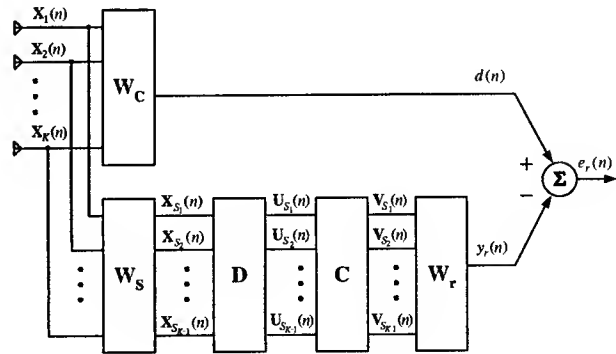


Figure 4: Block diagram of GSC using CAA algorithm.

$d(n)$ . The reduced rank signal vector,  $\mathbf{V}_S(n)$ , of the proposed GSC can be expressed using  $\mathbf{U}_S(n)$  and the rank-reducing matrix  $\mathbf{C}$  formed by the CAA algorithm as follows:

$$\mathbf{V}_S(n) = \mathbf{C}\mathbf{U}_S(n). \quad (12)$$

The auto-correlation matrix of  $\mathbf{V}_S(n)$  and the cross-correlation matrix between  $\mathbf{V}_S(n)$  and  $d(n)$  are as follows:

$$\mathbf{R}_{\mathbf{V}_S} = E[\mathbf{V}_S(n)\mathbf{V}_S^H(n)] = \mathbf{C}^H \mathbf{R}_{\mathbf{U}_S} \mathbf{C} \quad (13)$$

$$\mathbf{P}_{\mathbf{V}_S} = E[\mathbf{V}_S(n)d^*(n)] = \mathbf{C}^H \mathbf{P}_{\mathbf{U}_S} \quad (14)$$

The optimum weights of the proposed GSC can be expressed as

$$\mathbf{W}_{r,opt} = \mathbf{R}_{\mathbf{V}_S}^{-1} \mathbf{P}_{\mathbf{V}_S}. \quad (15)$$

Then the MSE of the proposed GSC can be given by

$$\begin{aligned} \zeta_r &= \sigma_d^2 - (\mathbf{C}^H \mathbf{P}_{\mathbf{U}_S})^H (\mathbf{C}^H \mathbf{R}_{\mathbf{U}_S} \mathbf{C})^{-1} (\mathbf{C}^H \mathbf{P}_{\mathbf{U}_S}) \\ &= \sigma_d^2 - \mathbf{P}_{\mathbf{V}_S}^H \mathbf{R}_{\mathbf{V}_S}^{-1} \mathbf{P}_{\mathbf{V}_S}. \end{aligned} \quad (16)$$

Since the proposed GSC minimizes any additional MSE caused by the partial adaptation, the MSE of the proposed GSC,  $\zeta_r$ , is almost equal to that of the full rank GSC,  $\zeta$ .

### 3.3. Censoring Spatial Smoothing Algorithm

Spatial smoothing can solve the signal cancellation phenomenon, however, it also requires high computation because it updates the weights corresponding to each subarray in each adaptation cycle [7]. Accordingly, this

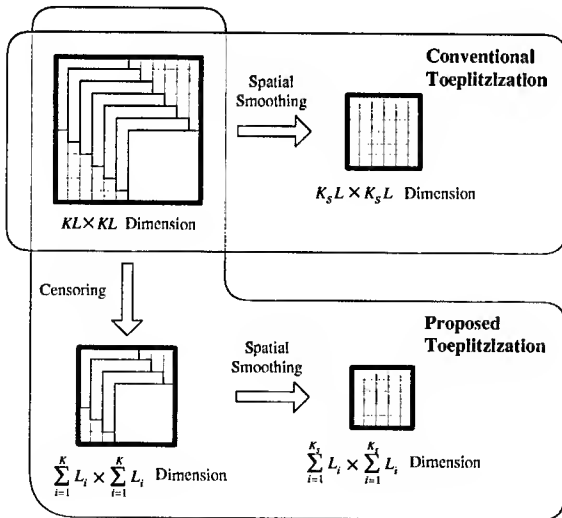


Figure 5: Conventional toeplitzization via spatial smoothing and proposed toeplitzization via censoring spatial smoothing.

paper proposes a censoring spatial smoothing, a combination of the CAA algorithm and spatial smoothing, which can solve both the signal cancellation phenomenon and the problem of the high computational complexity of frequency domain adaptive algorithms. Conventional toeplitzization sequence via spatial smoothing and the proposed toeplitzization sequence via censoring spatial smoothing are shown in Fig. 5.

## 4. SIMULATION RESULTS

To verify the performance of the proposed algorithm, several computer simulations were performed. The GSC was assumed to have 15 antenna elements which were divided into 7 subarrays with 9 antenna elements each. The length of the TDL was 8 and the initial false alarm rate used in the CAA algorithm was 1. One target signal with a Doppler frequency of 0.25Hz and a signal-to-noise power ratio (SNR) of 10dB was incoming from the broad side. Whereas three coherent interferences with a Doppler frequency of 0.25Hz and an interference-to-noise ratio (INR) of 40dB were incoming from  $-50^\circ$ ,  $-20^\circ$ , and  $34^\circ$ . The Doppler frequency was normalized with respect to the sampling frequency. The observation interval,  $N$ , to calculate the averaged output power,  $M_k$ , was set at 100.

Fig. 6 shows the learning curves of the GSC frequency domain and the proposed GSC using the CAA algorithm. Fig. 7 presents the simulation results of the proposed GSC including the variation in the false alarm rates, number of updated weights, and pattern response. The learning curves of the two GSCs considered were almost equal. However, after the learning curves were converged, the proposed GSC only adapted 8 weights, whereas the frequency domain GSC adapted

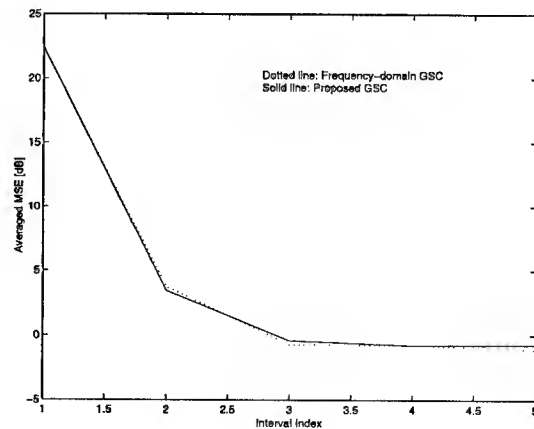
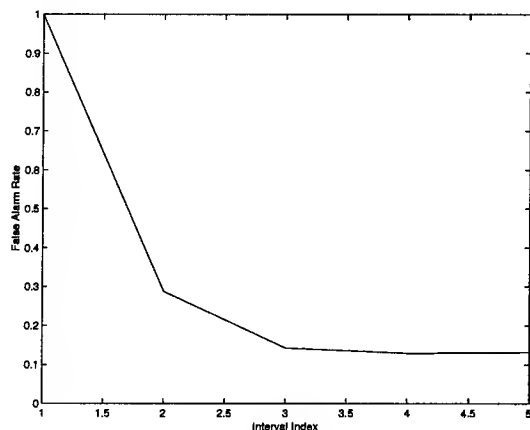
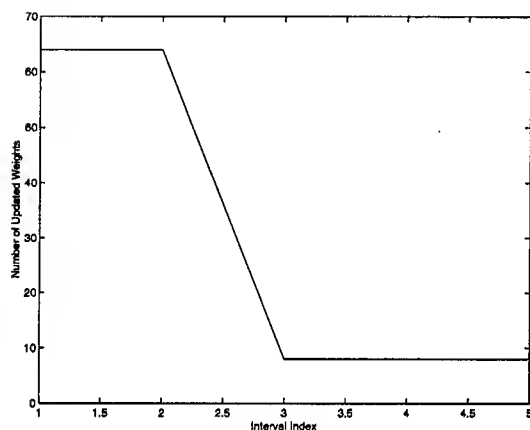


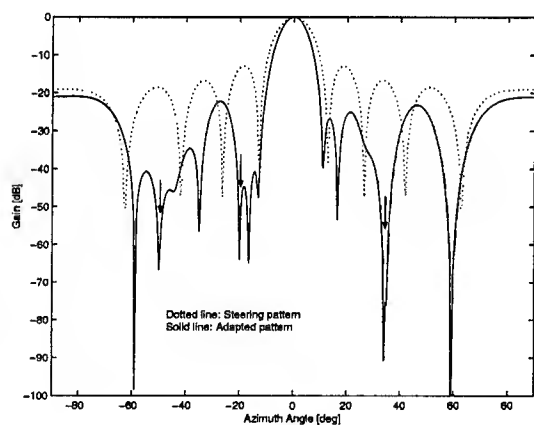
Figure 6: Learning curves of frequency domain GSC and proposed GSC.



(a)



(b)



(c)

Figure 7: Simulation results of proposed GSC: (a) Variation of false alarm rate, (b) Number of updated weights, (c) Steering pattern and Adapted pattern.

all 64 weights. Therefore, the required complex multiplication per adaptation cycle of the proposed GSC was 1,540 whereas that of the frequency domain GSC

was 2,912. The false alarm rate of the CA-CFAR processor was adaptively controlled and converged to an optimum value of about 0.13. The beam pattern shown in Fig 7(c) shows that deep nulls were formed in the adapted array pattern in the incident directions of every interference.

## 5. CONCLUSIONS

A new frequency domain partially adaptive algorithm, the CAA algorithm, was presented which can adaptively determine a subspace relative to the environment. In addition, a censoring spatial smoothing algorithm was proposed so that when combined with the CAA algorithm the computational complexity of the frequency domain adaptive algorithm was reduced plus the signal cancellation phenomenon was solved. Simulation results showed that the proposed GSC substantially reduces the computational complexity of the GSC frequency domain while maintaining the same level of performance.

## REFERENCES

- [1] B. Widrow, S. D. Stearns, *Adaptive Signal Processing*, Prentice Hall, Englewood Cliffs, NJ, 1985.
- [2] Y. H. Chen, H. D. Fang, "Frequency-domain Implementation of a Griffiths-Jim Adaptive Beamformer," *J. Acoust. Soc. Am.*, vol. 91, pp. 3354-3366, June 1992.
- [3] J. An, B. Champagne, "GSC Realisation Using the Two-dimensional Transform-domain LMS Algorithm," *IEE Proc. Radar, Sonar Navig.*, vol. 141, no. 5, pp. 270-278, Oct. 1994.
- [4] B. Widrow, et al, "Signal Cancellation Phenomena in Adaptive Antennas: Causes and Cures," *IEEE Trans. Antennas Propag.*, vol. 30, no. 3, pp. 427-445, July 1988.
- [5] P. P. Gandhi, S. A. Kassam, "Analysis of CFAR Processors in Nonhomogeneous Background," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 24, pp. 427-445, July 1988.
- [6] J. S. Goldstein, I. S. Reed, "Subspace Selection for Partially Adaptive Array Processing," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 33, no. 2, pp. 539-544, Apr. 1997.
- [7] T. J. Shan, T. Kailath, "Adaptive Beamforming for Coherent Signals and Interference," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 33, no. 3, pp. 527-536, June 1985.

# A NEW BEAMFORMING ALGORITHM BASED ON SIGNAL SUBSPACE EIGENVECTORS

*M. Biguesh<sup>1</sup>, S. Valaee<sup>2,1</sup>, B. Champagne<sup>3</sup>, M. H. Bastani<sup>1</sup>*

<sup>1</sup> Dept. of Elect Eng., Sharif University of Technology, Tehran, Iran

<sup>2</sup> Dept. of Elect Eng., Tarbiat Modares University, Tehran, Iran

<sup>3</sup> Dept. of Elect Eng., McGill University, Montreal, Québec, Canada

## ABSTRACT

A new beamforming algorithm, based on the eigendecomposition of the sample correlation matrix, has been introduced. The beamformer uses a weighted linear combination of the signal eigenvectors. Three versions of the beamformer have been proposed. It is shown that the proposed beamformer is a generalization of the delay-and-sum and the minimum variance beamformers. A linearly constrained minimum variance beamformer has also been derived. It is shown that the proposed approach induces robust beamformers.

## 1. INTRODUCTION

In various applications, one is concerned with extracting a desired signal immersed in noise and interference. Using an adaptive array, it is possible to avert the effect of interference and noise by an elaborate selection of array weights. Many algorithms maximize the array output signal to interference ratio (SINR) subject to knowing the direction of arrival (DOA) of the desired signal. In these cases, the weight vector is computed from the correlation matrix of interference and noise. We call this the signal-free correlation matrix (SFCM). However, if the desired signal DOA and the array geometry are known, one can use the correlation matrix of the received mixture of signal, noise, and interference, and attain the same result. Small errors in calibration and DOA estimation will cause signal cancellation [1, 2].

For most practical situations, noise and interference are mixed with signal, and the measurement of SFCM is not a simple task. To compute a signal-free correlation matrix, one can use the generalized sidelobe canceller (GSC) [3]. However, in this method, the calibration error or the desired signal DOA estimation error will cause a leakage of signal component which degrades the performance of method.

In many practical applications, the performance of detection depends on signal-to-interference ratio (SIR).

For instance, in spread spectrum communications, penetration of a smart jammer into the system, may cause a destructive effect on the system performance [4]. In such cases, interference minimization, rather than noise plus interference minimization, proves useful.

As a result of a higher resolution, much interest has been given to beamforming based on eigendecomposition [4, 5], and adaptive eigensubspace algorithms [6, 7]. Usually, these methods are based on the eigendecomposition of SFCM.

Here, we introduce a beamforming method based on the eigendecomposition of the received signal covariance matrix. To apply this beamforming method, one should know the DOA estimate of the desired signal, the number of point jammers, and an estimate of the received noise power. The introduced method, which needs a relatively low computation, is able to produce exact nulls in the direction of jammers. It is also able to maximize the output SINR or SNR. Due to lack of space, throughout, we omit the proof for the theorems.

## 2. SIGNAL MODEL

We assume an  $L$ -element array with arbitrary geometry and  $p$  narrowband point sources. Let  $\mathbf{x}(k)$  denote the complex data vector received by the array elements at the  $k$ 'th sampling instant. Data vector  $\mathbf{x}(k)$  can be expressed as a superposition of the received signals and noise as

$$\mathbf{x}(k) = \mathbf{A}(k)\mathbf{s}(k) + \mathbf{n}(k), \quad (1)$$

where  $\mathbf{n}(k)$  is the noise vector which is assumed to be white,  $\mathbf{s}(k)$  is the signal vector, and

$$\mathbf{A}(k) = [\mathbf{a}(\theta_1(k)), \dots, \mathbf{a}(\theta_p(k))], \quad (2)$$

with  $\mathbf{a}(\theta_i) = \mathbf{a}_i$  being the array steering vector at the direction  $\theta_i$ . Using (1), and assuming  $\sigma^2$  to be the noise power, the autocorrelation matrix of the received signal is obtained as

$$\mathbf{R}(k) = E\{\mathbf{x}(k)\mathbf{x}(k)^H\} = \mathbf{A}(k)\mathbf{\Gamma}\mathbf{A}(k)^H + \sigma^2\mathbf{I}, \quad (3)$$



where  $\mathbf{\Gamma} = \text{diag}(\gamma_1, \dots, \gamma_p)$  is the signal correlation matrix,  $E\{\cdot\}$  represents the expected value, and superscript  $H$  denotes Hermitian transposition. Diagonal form of  $\mathbf{\Gamma}$  is a consequence of the fact that the received signals are assumed to be uncorrelated with each other.

For the positive-definite correlation matrix  $\mathbf{R}$  one can find a set of eigenvalues  $(\lambda_i + \sigma^2)$ 's and orthonormal eigenvectors  $\mathbf{q}_i$ 's such that:

$$\mathbf{R}\mathbf{q}_i = (\lambda_i + \sigma^2)\mathbf{q}_i \quad \text{for} \quad 1 \leq i \leq L.$$

We assume that  $\lambda_i$ 's are in decreasing order, i.e.  $(\lambda_1 + \sigma^2) \geq \dots \geq (\lambda_L + \sigma^2)$ . It can be shown that  $\lambda_i = 0$  for  $i > p$ .

Eigenvectors  $\mathbf{Q} = [\mathbf{q}_1 \dots \mathbf{q}_L]$  can be divided into two matrices as  $\mathbf{Q} = [\mathbf{Q}_s | \mathbf{Q}_n]$  where the columns of  $\mathbf{Q}_s$  and  $\mathbf{Q}_n$ , respectively, span the orthogonal signal and noise subspaces. We can prove the following theorem.

**Theorem 1:** Defining  $\mathbf{\Lambda}_s = \text{diag}(\lambda_1 + \sigma^2 \dots \lambda_p + \sigma^2)$ , the following equalities are valid

$$\mathbf{\Lambda}_s^H \mathbf{Q}_s (\mathbf{\Lambda}_s - \sigma^2 \mathbf{I})^{-1} \mathbf{Q}_s^H \mathbf{A} = \mathbf{\Gamma}^{-1} \quad (4)$$

$$\mathbf{Q}_s^H \mathbf{A} \mathbf{\Gamma} \mathbf{A}^H \mathbf{Q}_s = (\mathbf{\Lambda}_s - \sigma^2 \mathbf{I}). \quad (5)$$

### 3. REDUCED-RANK BEAMFORMER

To extract the  $n$ 'th signal source (impinging on the array from direction  $\theta_n$ ), we propose the following beamforming weight vector

$$\mathbf{w}_{n,\epsilon} = \sum_{i=1}^p \frac{\mathbf{q}_i \mathbf{q}_i^H}{\epsilon \lambda_i + (1-\epsilon)\sigma^2} \mathbf{a}_n \quad \text{for} \quad 0 \leq \epsilon \leq 1. \quad (6)$$

This beamforming method, which needs the knowledge of the desired signal DOA and the number of point signal sources, has certain properties for various values of  $\epsilon$ . We study this beamforming method for three different values of  $\epsilon = 1, 0.5, 0$ , (noted by SC-1, SC-2 and SC-3, respectively).

#### 3.1. Special Case-1 (SC-1)

For this case, we compute the weight vector  $\mathbf{w}_n$  as

$$\mathbf{w}_n = \sum_{i=1}^p \frac{\mathbf{q}_i \mathbf{q}_i^H}{\lambda_i} \mathbf{a}_n = \mathbf{Q}_s (\mathbf{\Lambda}_s - \sigma^2 \mathbf{I})^{-1} \mathbf{Q}_s^H \mathbf{a}_n. \quad (7)$$

**Theorem 2:** The pattern for the SC-1 beamformer has null (exactly zero) in the direction of interferers, i.e.

$$\mathbf{w}_n^H \mathbf{a}_i = 0 \quad \text{for} \quad i = 1, \dots, p \quad \text{and} \quad i \neq n \quad (8)$$

**Theorem 3:** The output SINR of the SC-1 beamformer is restricted to  $\lambda_1/\sigma^2$  and  $\lambda_p/\sigma^2$ , i.e.

$$\frac{\lambda_p}{\sigma^2} \leq \left( \frac{S}{I+N} \right)_o \leq \frac{\lambda_1}{\sigma^2} \quad (9)$$

and for the case of only one signal source ( $p = 1$ ) the output SINR is equal to  $\lambda_1/\sigma^2$ .

#### 3.2. Special Case-2 (SC-2)

For this case, we compute the weight vector  $\mathbf{w}_n$  as

$$\mathbf{w}_n = \sum_{i=1}^p \frac{\mathbf{q}_i \mathbf{q}_i^H}{\lambda_i + \sigma^2} \mathbf{a}_n = \mathbf{Q}_s \mathbf{\Lambda}_s^{-1} \mathbf{Q}_s^H \mathbf{a}_n. \quad (10)$$

It can be shown that here,  $\mathbf{w}_n = \mathbf{R}^{-1} \mathbf{a}_n$ , which is the MV solution for the array weight vector — the MV beamformer maximizes the array output signal to interference and noise ratio (SINR) [5].

A shortcoming of the MV beamformer is its sensitivity to signal DOA uncertainty and array calibration error — this causes signal cancellation [8]. Define the output SINR sensitivity with respect to the steering vector error ( $\Delta \mathbf{a} = \tilde{\mathbf{a}} - \mathbf{a}$ ) as

$$S_{\text{SINR},\mathbf{a}}^{\mathbf{w}} = \frac{|\Delta \text{SINR}_o|}{\|\Delta \mathbf{a}\|^2}, \quad (11)$$

where

$$\Delta \text{SINR}_o = \text{SINR}_o|_{\tilde{\mathbf{a}}=\mathbf{a}+\Delta \mathbf{a}} - \text{SINR}_o|_{\tilde{\mathbf{a}}=\mathbf{a}}. \quad (12)$$

It can be shown that the SC-2 beamformer is less sensitive to the steering vector error (due to DOA uncertainty or uncalibrated array) when compared to the MV method — the sensitivity of MV beamformer increases rapidly with input SNR.

#### 3.3. Special Case-3 (SC-3)

For this reduced-rank beamformer, the weight vector,  $\mathbf{w}_n$ , is

$$\mathbf{w}_n = \sum_{i=1}^p \frac{\mathbf{q}_i \mathbf{q}_i^H}{\sigma^2} \mathbf{a}_n = \frac{1}{\sigma^2} \mathbf{Q}_s \mathbf{Q}_s^H \mathbf{a}_n. \quad (13)$$

Using  $\mathbf{Q}_s \mathbf{Q}_s^H = \mathbf{I} - \mathbf{Q}_n \mathbf{Q}_n^H$  and noting that the signal steering vector is orthogonal to the noise subspace, (13) can be written as

$$\mathbf{w}_n = \frac{1}{\sigma^2} \mathbf{a}_n. \quad (14)$$

If the true  $\mathbf{a}_n$  is known, the weight vector (14) produces the well-known delay-and-sum beamformer.

**Definition:** For an array with  $\mathbf{w}$  as a weight vector, we define the sensitivity of an array output SNR with respect to the array steering vector error ( $\Delta \mathbf{a} = \tilde{\mathbf{a}} - \mathbf{a}$ ) as

$$S_{SNR_o, \mathbf{a}}^{\mathbf{w}} = \frac{|\Delta SNR_o|}{\|\Delta \mathbf{a}\|^2} \quad (15)$$

where

$$\Delta SNR_o = SNR_o|_{\tilde{\mathbf{a}}=\mathbf{a}+\Delta \mathbf{a}} - SNR_o|_{\tilde{\mathbf{a}}=\mathbf{a}}. \quad (16)$$

It can be proved that the output SNR for the SC-3 beamformer is less sensitive to the array steering vector error than the delay-and-sum beamformer.

For SC-3, the array output SNR is

$$\frac{S_o}{N_o} = L \left( \frac{S}{N} \right)_i. \quad (17)$$

We have proved that the maximum output SNR for an array with  $L$  elements is  $L$  times the input SNR. Thus, SC-3 maximizes the output SNR.

### 3.4. Improved LCMV method

As mentioned earlier, the SC-2 beamformer has the properties of MV method with a smaller sensitivity. In the MV method, the weight vector is the solution of the following minimization

$$\min_{\mathbf{w}} \{\mathbf{w}^H \mathbf{R} \mathbf{w}\} \quad \text{subject to} \quad \mathbf{w}^H \mathbf{a}(\theta_i) = g \quad (18)$$

where  $g$  is a constant. By the method of Lagrange multipliers, the solution to this minimization is

$$\mathbf{w} = g \frac{\mathbf{R}^{-1} \mathbf{a}}{\mathbf{a}^H \mathbf{R}^{-1} \mathbf{a}}. \quad (19)$$

The single linear constraint in (18) can be generalized to a multiple linear constraint. For instance, to produce a beampattern with a unit gain in the direction of sources,  $\theta_1$  and  $\theta_2$ , the desired constraint may be expressed as

$$\begin{bmatrix} \mathbf{a}^H(\theta_1) \\ \mathbf{a}^H(\theta_2) \end{bmatrix} \mathbf{w} = \begin{bmatrix} 1 \\ 1 \end{bmatrix}. \quad (20)$$

If there are  $m < L$  linear constraints on  $\mathbf{w}$ , it is possible to write them in the matrix form  $\mathbf{C}^H \mathbf{w} = \mathbf{f}$ , where the  $L \times m$  matrix  $\mathbf{C}$  and the  $m$ -dimensional vector  $\mathbf{f}$  are the constraint matrix and the response vector, respectively. It is assumed that the constraints are linearly independent — the constraint matrix has rank  $m$ . The solution of (18) is then

$$\mathbf{w} = \mathbf{R}^{-1} \mathbf{C} [\mathbf{C}^H \mathbf{R}^{-1} \mathbf{C}]^{-1} \mathbf{f} \quad (21)$$

which is called the linear constraint minimum variance (LCMV) weight vector. Similar to (6), we define the following weight vector which satisfies the constraint  $\mathbf{C}^H \mathbf{w} = \mathbf{f}$ ,

$$\mathbf{w} = \mathbf{H} \mathbf{C} [\mathbf{C}^H \mathbf{H} \mathbf{C}]^{-1} \mathbf{f}, \quad (22)$$

where  $\mathbf{H}$  is defined as

$$\mathbf{H} \stackrel{\text{def}}{=} \sum_{i=1}^p \frac{\mathbf{q}_i \mathbf{q}_i^H}{\epsilon \lambda_i + (1 - \epsilon) \sigma^2} \bigg|_{\epsilon=0.5}. \quad (23)$$

We call this technique, the improved LCMV (ILCMV) algorithm. Replacing Karhunen-Loeve expansion in (21), it is straightforward to prove that when the columns of  $\mathbf{C}$  are a subset of the columns of  $\mathbf{A}$ , the weight vector (21) is the same as  $\mathbf{w}$  in (22). However, the simulation results show an improvement for ILCMV when compared to LCMV.

## 4. SIMULATION RESULTS

In the following examples, we use a uniform circular array (UCA) with  $L$  omnidirectional antenna elements. The interelement spacing is assumed to be  $\lambda/2$  where  $\lambda$  is the received signal wavelength. Three stationary point signal sources with the same power are used in simulations.

The effect of  $\epsilon$  in (6) on the produced pattern for  $\epsilon = 0, 0.2, 0.4, 0.6, 0.8$  is shown in Fig. 1, for the desired source at  $180^\circ$ , and interfering sources at  $125^\circ$ , and  $280^\circ$  ( $L = 8$ ). The figure shows that as  $\epsilon$  increases, the two relative nulls of the beampattern move towards the jammers and become deeper.

In the second example, we choose a random DOA for signal and jammers. Fig. 2 shows the average output SIR, SINR, and SNR as a function of  $\epsilon$  for the proposed beamformer choosing a random DOA for signals. The input SNR is assumed to be 3dB and  $L = 8$  is considered. The curves show that the output SIR decreases rapidly with decreasing  $\epsilon$ , however, the changes in SNR and SINR are not substantial.

In Fig. 3, the produced beampatterns with LCMV and ILCMV methods are compared (for  $L = 15$ ). Here, the signal source DOAs are at  $80^\circ$  and  $240^\circ$ , and an interferer is located at  $160^\circ$ . The results clearly show the robustness of the proposed method against DOA uncertainty and array calibration error.

## REFERENCES

- [1] B. Friedlander, "A Signal Subspace Method for Adaptive Interference Cancellation," *IEEE Trans. Acoust., Speech and Sig. Proc.*, Vol. ASSP-36, No. 12, pp. 1853-1845, Dec. 1988.

- [2] A. P. Applebaum and D. J. Chapman, "Adaptive Array with Main Beam Constraints," *IEEE Trans. Ant. and Prop.*, Vol. AP-24, No. 5, pp. 650-662, Sep 1976.
- [3] A. Farina, "Antenna-Based Signal Processing Techniques for Radar Systems." Artech House, 1992.
- [4] A. M. Haimovich and Y. Bar-Ness, "An Eigenanalysis Interference Canceller," *IEEE Trans. on Sig. Proc.*, Vol. SP-39, No. 1, pp. 76-84, Jan. 1991.
- [5] P. A. Zulch, et al, "Comparison of Reduced-rank Signal Processing Techniques," 32nd Asilomar Conference Signal System and Computer, pp. 421-425, 1998.
- [6] B. Champagne, "Adaptive Eigendecomposition of Data Covariance Matrices Based on First-Order Perturbation," *IEEE Trans. on Sig. Proc.*, Vol. SP-42, No. 10, pp. 2758-2770, Oct. 1994.
- [7] J. F. Yang and M. Kaveh, "Adaptive Eigensubspace Algorithms for Direction or Frequency Estimation and Tracking," *IEEE Trans. Acoust., Speech and Sig. Proc.*, Vol. ASSP-36, No. 2, pp. 241-251, Feb. 1988.
- [8] H. Cox, R. M. Zeskind, and M. M. Owen, "Robust Adaptive Beamforming," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-35, pp. 1365-1378, Oct. 1987.

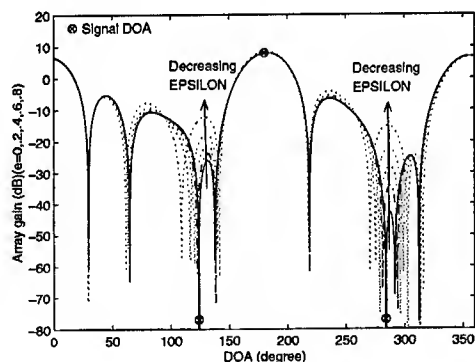


Figure 1: The effect of  $\epsilon$  on the produced pattern for  $\epsilon = 0, 0.2, 0.4, 0.6, 0.8$ . The desired source is located at  $180^\circ$ , and interfering sources are at  $125^\circ$ , and  $280^\circ$  ( $L = 8$ ).

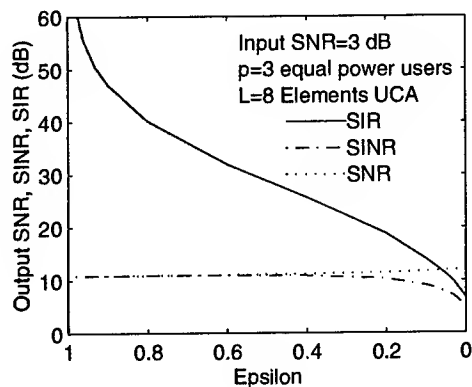


Figure 2: The average output SIR, SINR, and SNR as a function of  $\epsilon$  for an 8-element UCA.

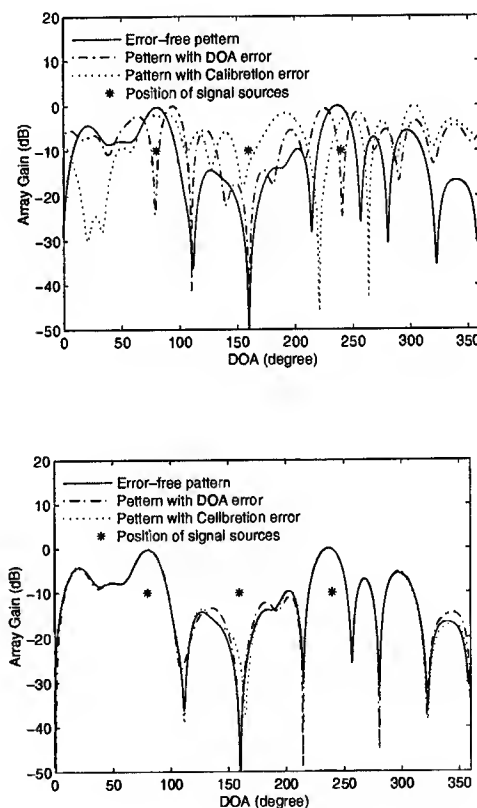


Figure 3: Beampattern for the LCMV (top) and the ILCMV (bottom) algorithms.

# DETECTION OF SOURCES IN ARRAY PROCESSING USING THE BOOTSTRAP

Ramon Brcich, Per Pelin and Abdelhak Zoubir

Australian Telecommunications Research Institute (ATRI)\* &  
School of Electrical and Computer Engineering,  
Curtin University of Technology, GPO Box U1987, Perth WA 6845, Australia.  
Email : r.brcich@ieee.org

## ABSTRACT

A hypothesis testing methodology for determining the number of narrowband sources impinging on an array is presented. Using multiple hypothesis tests the multiplicity of the smallest ordered eigenvalues of the sample correlation matrix and hence the number of sources, is determined. The finite sample null distributions of the test statistics are estimated using bootstrap resampling. By removing the assumption of Gaussianity and large sample size that the traditional MDL approach is based on, we are able to gain improvements in the small sample case or when there are deviations from Gaussianity.

## 1. INTRODUCTION

The first step in most array signal processing problems is to determine the number of narrowband sources impinging on an array. Traditional approaches are based on the application of information theoretic criteria, such as Rissanen's Minimum Description Length (MDL), to an estimate of the likelihood function of the ordered eigenvalues of the data [8, 5]. For Gaussian data, the MDL is asymptotically consistent and becomes a simple function of the ordered sample eigenvalues.

Instead of using the MDL, whose behaviour is uncertain for small sample sizes or non-Gaussian data, we estimate the small sample distributions of the ordered eigenvalues using a bootstrap resampling technique [4]. A multiple hypothesis test is then applied, sequentially testing for equality of the smallest ordered eigenvalues to determine the number of sources.

By estimating the distributions of the ordered eigenvalues, detection rates can be improved when the sample size is small, or when the signal deviates from Gaussianity.

The paper is organised as follows. In section 2 the signal model is described before discussing the testing methodology and the use of multiple hypothesis tests in section 3. In section 4 the use of the bootstrap in estimating the null distributions of the test statistics is explained. Section 5 points out the need to reduce the bias of the sample ordered eigenvalues and describes the method of jackknife bias reduction. Finally, section 6 compares the proposed method against the MDL, followed by some conclusions.

\*This work was in part supported by the Australian Telecommunications Cooperative Research Centre (AT-CRC).

## 2. SIGNAL MODEL

We receive  $n$  i.i.d snapshots,  $\mathbf{x}(t)$ , of complex data from a  $p$  element array,

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{v}(t), \quad t = 1, \dots, n \quad (1)$$

where  $\mathbf{A}$  is the  $p \times q$  array steering matrix,  $\mathbf{s}(t)$  is the  $q$  ( $q < p$ ) vector valued source signal and  $\mathbf{v}(t)$  is spatially white additive noise with variance  $\sigma^2$ .

The correlation matrix of the snapshots is then

$$\mathbf{R} = \mathbb{E}[\mathbf{x}(t)\mathbf{x}^H(t)] = \mathbf{A}\mathbf{R}_s\mathbf{A}^H + \sigma^2\mathbf{I} \quad (2)$$

where  $\mathbf{R}_s = \mathbb{E}[\mathbf{s}(t)\mathbf{s}^H(t)]$ . Let the ordered eigenvalues of  $\mathbf{R}$  be  $\lambda_1 \geq \dots \geq \lambda_q > \lambda_{q+1} = \dots = \lambda_p$ . This suggests we estimate  $q$  by determining the multiplicity of the smallest ordered eigenvalues of the sample correlation matrix,

$$\hat{\mathbf{R}} = \frac{1}{n-1} \sum_{t=1}^n \mathbf{x}(t)\mathbf{x}^H(t). \quad (3)$$

## 3. HYPOTHESIS TESTING

To test for multiplicity of the smallest ordered eigenvalues we consider the following situations where the corresponding number of sources,  $q$ , is as stated,

$$\begin{array}{llllll} \lambda_1 & = & \lambda_2 & = & \dots & = & \lambda_p & q & = & 0 \\ & & \lambda_2 & = & \dots & = & \lambda_p & q & = & 1 \\ & & & & \ddots & & \vdots & & & \\ & & & & & & \lambda_{p-1} & = & \lambda_p & q & = & p-2 \\ \lambda_1 & \neq & \lambda_2 & \neq & \dots & \neq & \lambda_p & q & \geq & p-1 \end{array}$$

To determine  $q$  we are required to test which of the above conditions is true. To accomplish this we propose the following procedure,

1. Set  $k = 0$ .
2. Test for equality of the smallest  $p - k$  eigenvalues. If this hypothesis is accepted then  $\hat{q} = k$  and the procedure is finished.
3. If we rejected the hypothesis and  $k < p - 2$  then set  $k = k + 1$  and return to step 2. If  $k = p - 2$  then  $\hat{q} = p - 2$  was rejected, so we must assume all the eigenvalues are unequal and  $\hat{q} \geq p - 1$ .

To perform the joint test for equality of the smallest eigenvalues at each stage of the procedure we make use of Roy's union intersection (UI) method. Roy's UI method allows us to construct a test for any joint hypothesis,  $H$ , providing it can be expressed as an intersection of simpler subhypotheses,  $H_i$  for which tests exist. That is, if  $H = \cap_i H_i$  and tests for the  $H_i$  are available, then the rejection region for  $H$  is given by the union of rejection regions of the  $H_i$ . This means the global null hypothesis,  $H$ , is rejected if at least one of the  $H_i$  is rejected. From this we can define the family wise error rate (FWE) as

$$\text{FWE} = \Pr(\text{Reject at least one } H_i \mid \text{all } H_i \text{ are true}).$$

The FWE plays a similar role to that of the set level,  $\alpha$ , in univariate testing.

Following the UI principle a test for  $\lambda_k = \dots = \lambda_p$  can be constructed from the hypotheses  $H_{ij} : \lambda_i = \lambda_j$  where the rejection region is given by  $\cap_{i=k, j>i}^{p-1} H_{ij}$ . The special case of  $\lambda_1 \neq \dots \neq \lambda_p$  is chosen when all other hypotheses have been rejected, as already stated. Note that the alternative hypothesis to each of the  $H_{ij}$  is  $K_{ij} : \lambda_i > \lambda_j$ .

To test each of the  $H_{ij}$  while maintaining the FWE we must use a multiple test procedure. Here, both Bonferonni's single step and Holm's sequentially rejective Bonferonni algorithm (SRB) are used.

The Bonferonni method tests each of the  $H_{ij}$  at a level of  $\alpha/l$  where  $l$  is the number of hypothesis being tested. Assuming the significance levels are independent and uniformly distributed on  $[0, 1]$  this method exactly controls the FWE at level  $\alpha$ .

In this problem not all the hypotheses are independent. There are logical implications between hypotheses so that the truth/falsehood of some imply the truth/falsehood of others. For instance, if  $H_{1p}$  were true, then this would imply all the  $H_{ij}$  were true.

When the hypotheses implicate each other, stepwise methods such as Holm's SRB strongly control the FWE. For more details on Holm's SRB and multiple hypothesis testing in general see [9].

#### 4. BOOTSTRAP PROCEDURE

To evaluate the significance levels for the multiple hypothesis tests we require the null distribution of each test statistic,  $T_{ij} = \hat{\lambda}_i - \hat{\lambda}_j$ , where  $i, j$  are defined as for  $H_{ij}$ .

We use the bootstrap [4], to estimate the null distributions. Briefly, we randomly resample from the matrix of array snapshots  $\mathbf{X} = (\mathbf{x}(1), \mathbf{x}(2), \dots, \mathbf{x}(n))$  to generate a bootstrap data set,  $\mathbf{X}^*$ . Recalculating the test statistic from  $\mathbf{X}^*$  gives us  $T_{ij}^* = \hat{\lambda}_i^* - \hat{\lambda}_j^*$ , the bootstrap procedure for resampling eigenvalues is summarised in Table 1. Repeating this procedure  $B$  times gives us the set of bootstrapped test statistics,  $T_{ij}^*(b)$ ,  $b = 1, \dots, B$ .

Critical points or significance levels can then be found from the test statistic,  $T_{ij}$ , and the bootstrap distribution,  $T_{ij}^*$ , by forming  $T_{ij}^*(b) - T_{ij}$ , which approximates the distribution of  $T_{ij}$  under the null [4].

In [1, 2] the performance of the bootstrapped eigenvalues is considered. They show that the bootstrap converges to the asymptotic distributions for distinct eigenvalues, but not for multiple eigenvalues. This may be attributed to the

Table 1: Bootstrap procedure for resampling eigenvalues.

1. Randomly select a single snapshot from the matrix of array snapshots with replacement (a column of  $\mathbf{X}$ ).
2. Repeat the random selection  $n$  times to obtain a resample of the matrix of array snapshots,  $\mathbf{X}^*$ .
3. Estimate the sample correlation matrix of  $\mathbf{X}^*$ ,  $\hat{\mathbf{R}}^*$ .
4. Centre  $\mathbf{X}^*$  by subtracting the column-wise mean from each column.
5. The resampled eigenvalues,  $\hat{\lambda}_1^*, \dots, \hat{\lambda}_p^*$  are estimated from the centred  $\hat{\mathbf{R}}^*$ .
6. Repeat steps 1 to 5,  $B$  times to obtain the bootstrap set of eigenvalues  $\hat{\lambda}_1^*(b), \dots, \hat{\lambda}_p^*(b)$ ,  $b = 1, \dots, B$ .

complex nonlinear nature of eigenvalue estimation. However, it is shown that bootstrapping with fewer resamples,  $m < n$ , where  $\min(m, n) \rightarrow \infty$  and  $m/n \rightarrow 0$ , ensures the bootstrap converges to the asymptotic distribution.

For the small sample sizes considered here it is difficult to fulfill these conditions without increasing the error in the bootstrap distribution due to the reduced number of resamples adversely affecting the bias and variance of the resampled eigenvalues. However, for large sample sizes we may reduce the error more effectively by subsampling (resampling  $m < n$  times without replacement) and estimating the rate of convergence of the distribution to its asymptotic limit. The conditions under which subsampling is valid encompass a wider range of distributions and statistics than the bootstrap, more on the theory of subsampling and rate estimation may be found in the recent publication [3].

Although we have investigated the use of subsampling for this problem, the bootstrap was found to provide a sufficiently accurate estimate for the eigenvalue distributions considering the sample sizes used.

#### 5. BIAS REDUCTION

Though the sample eigenvalues are asymptotically unbiased, the bias may be significant for finite sample sizes. Here we are particularly concerned with small sample performance and some bias reduction is required. The reason we need bias reduction stems from the nature of the hypothesis tests and the use of the bootstrap. A bias in the estimator will remain in the test statistic but will be essentially removed from the bootstrap estimate of the null distribution. This occurs as the estimate of the null distribution is created by subtracting the test statistic from the bootstrapped test statistics, as the bias of the test statistic is very near the bias of the bootstrapped test statistics the bias essentially subtracts out.

The expected value of the  $q$  distinct sample eigenvalues is [6]

$$E[\hat{\lambda}_i] = \lambda_i + \frac{\lambda_i}{n} \sum_{j=1, j \neq i}^p \frac{\lambda_j}{\lambda_i - \lambda_j} + O\left(\frac{1}{n^2}\right). \quad (4)$$

Subtracting the term in  $1/n$  from  $\hat{\lambda}_i$  and replacing the true eigenvalues by their estimates gives the bias reduced estimates of the  $q$  distinct eigenvalues,

$$\hat{\lambda}_i^c = \hat{\lambda}_i \left( 1 - \frac{1}{n} \sum_{j=1, j \neq i}^q \frac{\hat{\lambda}_j}{\hat{\lambda}_i - \hat{\lambda}_j} - \frac{p-q}{n} \frac{\lambda}{\hat{\lambda}_i - \lambda} \right), \quad i = 1, \dots, q, \quad (5)$$

where  $\lambda$  is the value of the multiple eigenvalues,  $\lambda = \lambda_{q+1} = \dots = \lambda_p$ , which may be replaced with the maximum likelihood estimate,

$$\hat{\lambda}_i^c = \frac{1}{p-q} \sum_{j=q+1}^p \hat{\lambda}_j, \quad i = q+1, \dots, p. \quad (6)$$

The corrected eigenvalues of (5) have a bias of order  $1/n^2$ , while those of (6) are unbiased.

It is not possible to use these bias reduced estimates blindly as the multiplicity of the eigenvalues is required. Also, the difference between successive distinct eigenvalues must be large compared to the sampling errors, which are of order  $1/\sqrt{n}$ . If this condition is not fulfilled, the variance of the corrected distinct eigenvalues can increase dramatically [6]. This is easily understood by considering the effect of very close distinct eigenvalues on the denominator of the summation in (5).

We evaluated several alternative techniques for bias reduction based on resampling methods. The advantage of resampling techniques to bias reduction in our case is that they may be applied blindly, with no knowledge of the eigenvalue multiplicity. The jackknife was found to be most effective scheme, it reduced the bias at least as much as (5) and did not suffer from any large increases in variance, even for multiple eigenvalues.

We must also consider the effects of non-Gaussianity on the bias. Here the jackknife has an advantage over (5) which was derived under the assumption of Gaussianity. For non-Gaussian data the bias also depends on the cumulants of the underlying distribution [7]. In the non-Gaussian case then, the jackknife is still valid as it is a distribution free, though not distribution insensitive, method.

The procedure for jackknife bias reduction is given in Table 2, more details may be found in [4]. In all cases jackknife bias reduction was applied to eigenvalue estimates and bootstrapped eigenvalues. Applying bias reduction to the bootstrapped eigenvalues is necessary as it alters the variance of the estimate and this change in the test statistic must be matched in the estimate of the null distribution. It also helps to mitigate any residual bias in estimating the null distribution.

## 6. SIMULATIONS

In the following simulations the proposed method is evaluated by comparing it to the MDL [8] in a variety of scenarios. Both the Bonferonni procedure and Holm's SRB are shown. Some parameters which remain unchanged throughout the tests are: the number of resamples,  $B = 200$ , the FWE,  $\alpha = 0.02$  and the element spacing which was one half the wavelength. The signals are also Gaussian, unless

Table 2: Jackknife Bias Reduction

1. Given the matrix of array snapshots  $\mathbf{X}$ , define the  $i^{\text{th}}$  jackknife sample of  $\mathbf{X}$  to be  $\mathbf{X}_{(i)} = (\mathbf{x}(1), \dots, \mathbf{x}(i-1), \mathbf{x}(i+1), \dots, \mathbf{x}(n))$ .
2. Let  $\hat{\lambda}_1^{(i)}, \dots, \hat{\lambda}_p^{(i)}$  be the ordered eigenvalues estimated from  $\mathbf{X}_{(i)}$ .
3. Compute  $\hat{\lambda}_j^{(\cdot)} = 1/n \sum_{i=1}^n \hat{\lambda}_j^{(i)}$ .
4. The bias reduced eigenvalues are given as  $\hat{\lambda}_j^c = \hat{\lambda}_j - (n-1)(\hat{\lambda}_j^{(\cdot)} - \hat{\lambda}_j)$ .

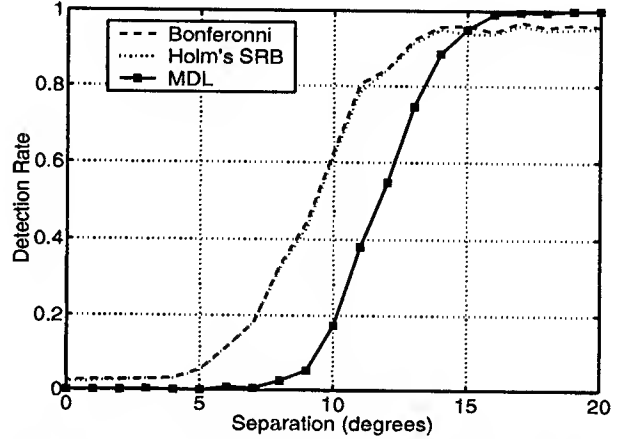


Figure 1: Empirical probability of correctly detecting two narrowly separated sources as the direction of one is varied.

otherwise stated. All results were averaged over 400 Monte Carlo simulations.

**Angular resolution :** (Figure 1) We have a  $p = 4$  element array with  $q = 2$  sources. The first source is fixed at 20 degrees (with respect to broadside) while the other is allowed to vary between 20 and 40 degrees. Both sources were at 0dB SNR and there were  $n = 100$  snapshots.

**Effect of SNR :** (Figure 2) The conditions are the same as above except that the second source was removed and the SNR was varied.

**Correlated sources :** (Figure 3) We have a  $p = 6$  element array with  $q = 2$  correlated sources at 20 and 40 degrees and SNR's of -3 and 0dB respectively. The correlation coefficient between the two sources is varied from 0.66 to 0.99.

While the MDL appears superior under the ideal conditions of widely separated sources, high SNR or weakly correlated sources there is a noticeable improvement for the more difficult cases of narrowly separated sources, low SNR and highly correlated sources. For example, at an SNR of -7dB, the proposed method correctly detects the single source at a rate of 80% while the MDL is at 40%. One point to note is that both Bonferonni and Holm's methods behave very similarly, this is commented on later.

**Sample Size :** (Figure 4) We have a  $p = 4$  element array with  $q = 1$  source at 20 degrees and -7dB SNR. The sample size was varied over  $10 \leq n \leq 250$ .

As suspected we notice an improvement in the small

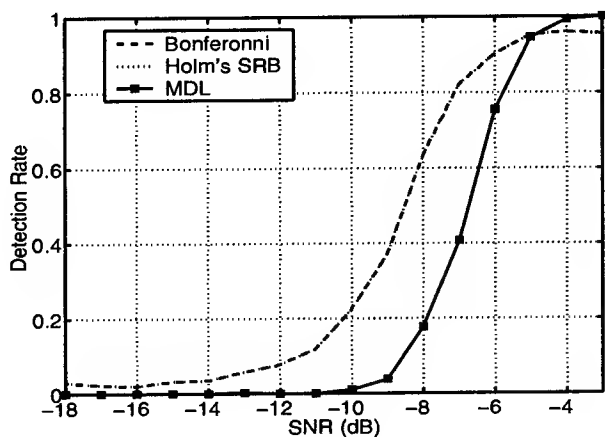


Figure 2: Empirical probability of correctly detecting a single source as the SNR is varied.

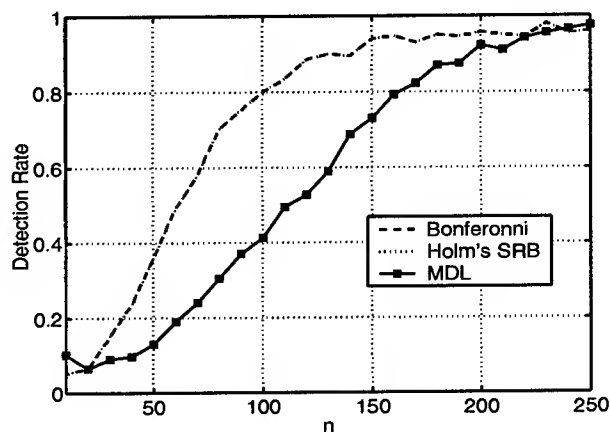


Figure 4: Empirical probability of correctly detecting a single source as the sample size is varied.

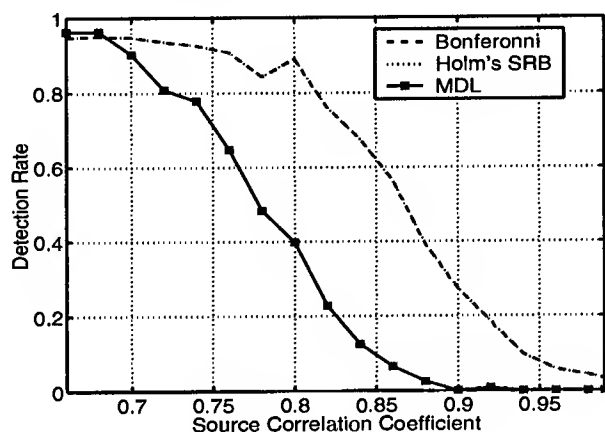


Figure 3: Empirical probability of correctly detecting two correlated sources as the correlation coefficient is varied.

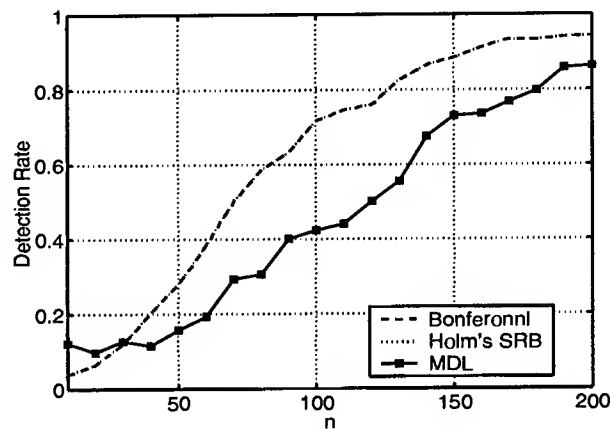


Figure 5: Empirical probability of correctly detecting a single Laplacian source in Gaussian noise as the sample size varies.

sample case up to  $n = 250$ , suggesting there is an advantage in using the proposed method for small sample sizes. For instance, with  $n = 100$  the detection rates are 80% and 40% for the proposed method and MDL respectively.

**Non-Gaussianity :** (Figures 5, 6) Here we have the same conditions as above, except that for Figure 5 we have a single Laplacian source in Gaussian noise while for Figure 6 we have a single Gaussian source in Laplacian noise. The sample size was varied over  $10 \leq n \leq 200$ .

Comparing both these non-Gaussian cases to the previous Gaussian example it is apparent that the behaviour of the proposed methods with respect to sample size is similar. Detection rates do drop for the non-Gaussian cases, though the improvement over the MDL is clear, suggesting the method is more robust than the MDL to deviations from Gaussianity.

**FWE :** (Figure 7) Here we show the probability of correctly accepting the global null hypothesis, that all eigenvalues are equal, for a  $p = 4$  element array with  $q = 0$  sources. The sample size was varied over  $10 \leq n \leq 250$ .

In a source free environment the FWE rate, or the probability of rejecting the null hypothesis that all eigenvalues

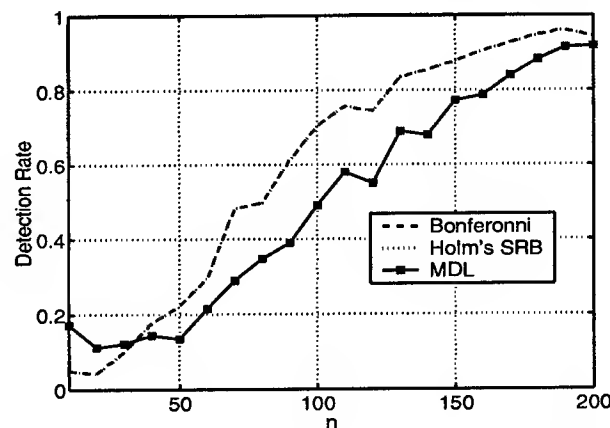


Figure 6: Empirical probability of correctly detecting a single Gaussian source in Laplacian noise as the sample size varies.

are equal, should be maintained close to the set level, which in this case is  $\alpha = 0.02$ . It can be seen that the FWE is not

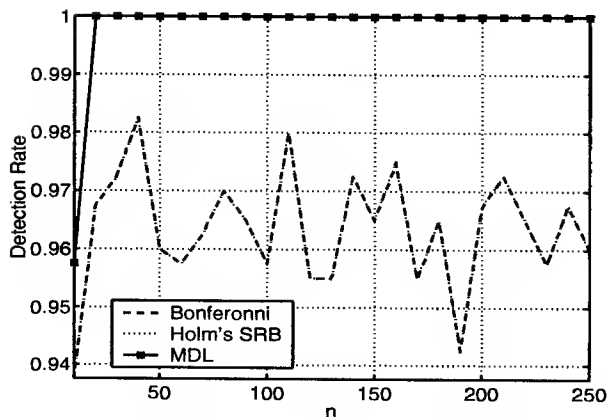


Figure 7: Empirical probability of correctly identifying  $q = 0$  sources in a source free environment as the sample size varies.

exactly maintained, instead it is approximately 0.03. Investigation of the significance levels showed they are slightly nonuniform, approximately 3% were less than 0.02. This can be attributed to the errors in estimating the null distribution of the hypotheses. Thus the assumption of uniform significance levels in the Bonferroni methods is not true and the FWE cannot be exactly maintained, explaining the attained FWE near 0.03. Disregarding variation due to the finite number of Monte Carlo realisations used, the level of the test appears to be constant with respect to sample size. This is expected as the level of the test should be independent of the array size and the sample size. The FWE should be kept small ( $< 0.05$ ) to avoid unnecessary false detections, however, as the FWE decreases we must take more resamples to properly estimate the critical points and the computational load increases [4]. Also, decreasing the FWE decreases the power of the test and the performance improvement over the MDL will decrease.

**Source saturated environment :** Finally we show an example when we have  $q = 3$  sources and a  $p = 4$  element array. This is the maximum number of sources we can detect for  $p = 4$ . The sources were at 10, 30 and 50 degrees with SNR's of -2, 2 and 6dB respectively. The sample size was varied over  $10 \leq n \leq 200$ .

Here we can see a large improvement over the MDL, suggesting the proposed method is well suited to source saturated environments.

In general it appears that the proposed method is relatively insensitive to the multiple test procedure. It is difficult to determine the cause/s of this, though error in estimating the null distribution is certainly important. We are currently investigating whether more powerful tests specifically tailored to the implications among the hypotheses will yield an improvement.

## 7. CONCLUSION

Here we approached the source detection problem in array processing from a hypothesis testing viewpoint. Instead of using information theoretic criteria designed for large samples and Gaussian signals, we test for equality of the

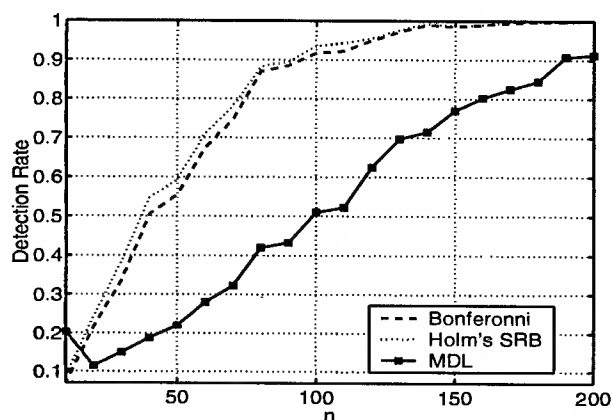


Figure 8: Empirical probability of correctly detecting  $q = p - 1 = 3$  sources as the sample size varies.

smallest ordered eigenvalues of the sample correlation matrix using multiple hypothesis tests. By estimating the finite sample null distributions of the test statistics using the bootstrap we show an improvement over the MDL for small sample sizes or when there are deviations from Gaussianity. The proposed method also performs favorably compared to the MDL under a variety of situations including low SNR, strong source correlation, narrowly separated sources and saturated source environments.

## REFERENCES

- [1] R. Beran and M. Srivastava. Bootstrap tests and confidence regions for functions of a covariance matrix. *The Annals of Statistics*, 13(1):95–115, 1985.
- [2] R. Beran and M. Srivastava. Correction : Bootstrap tests and confidence regions for functions of a covariance matrix. *The Annals of Statistics*, 15(1):470–471, 1987.
- [3] J. Romano D. Politis and M. Wolf. *Subsampling*. Springer-Verlag, 1999.
- [4] B. Efron and R. Tibshirani. *An Introduction to the Bootstrap*. Chapman and Hall, 1993.
- [5] E. Fishler and H. Messer. Order statistics approach for determining the number of sources using an array of sensors. *IEEE Signal Processing Letters*, 6(7):179–82, July 1999.
- [6] D. Lawley. Tests of significance for the latent roots of covariance and correlation matrices. *Biometrika*, 43:128–136, 1956.
- [7] C. Waternaux. Asymptotic distribution of the sample roots for a nonnormal population. *Biometrika*, 63(3):639–45, 1976.
- [8] M. Wax and T. Kailath. Detection of signals by information theoretic criteria. *IEEE Transactions on Acoustics, Speech and Signal Processing*, ASSP-33(2):387–92, April 1985.
- [9] P. Westfall and S. Young. *Resampling-Based Multiple Testing*. John Wiley & Sons, 1993.



# ROBUST LOCALIZATION OF SCATTERED SOURCES

*Joseph Tabrikian*

Dept. of ECE  
Ben-Gurion University  
Beer Sheva 84105, ISRAEL

*Hagit Messer*

Dept. of EE - Systems  
Tel Aviv University  
Tel Aviv 69978, ISRAEL

## ABSTRACT

This paper presents a new robust algorithm for scattered source localization. The proposed algorithm is based on a decomposition of the channel vector into subspaces characterized by their sensitivities to the spatial source parameters, such as the source spread which is usually treated as an unknown nuisance parameter. This decomposition isolates a subspace of the data which is not a function of the unknown nuisance parameters, and the resulting estimator does not involve any search over these parameters. The Maximum-Likelihood estimator for the new decomposed model is developed. The estimator uses only the information carried by the insensitive subspace of the data while perturbations of the channel vector in the sensitive subspace are assumed to be unknown parameters. Identification of the insensitive subspace is done according to the channel vector covariance matrix. Simulation results are presented to demonstrate the effectiveness of the proposed algorithm.

## 1. INTRODUCTION

Traditional array processing techniques assume a wave-field generated by point sources. However, point source assumption does not hold in many practical problems. For instance, multipath propagation in mobile radio communications or low-elevation radio link affect the spatial distribution of the observed signal. It may be more reasonable to assume that most of the energy incident on the array is from local scattering near the transmitter. The diffuse propagation can then be described by the superposition of a large number of plane waves, reflected by so-called reflectors standing for as much point sources and distributed around the real direction of the transmitter.

In their pioneering work, Valaee *et al* [1] presented the problem of distributed sources and proposed a parametric approach to localize distributed sources. They consider incoherently and coherently distributed sources where the angular signal density is assumed to be known. In [2] and [3] a generalized array manifold model was

used for DOA estimation in channel with local scattering assuming fully coherent reflections in which it was shown that local scattering, has significant impact on direction-of-arrival (DOA) estimation for time-invariant channels. Performance limits of distributed source localization has been studied through Cramer-Rao bound (CRB). A partially coherent distributed (PCD) concept was considered in [4]. In [5], performance of the method of [1] is studied through the CRB for fully coherent and incoherent distributed sources.

The problem of distributed sources has been investigated through parametric models in which the spatial source parameters are unknown. These parameters consist of parameters of interest such as the nominal DOA, as well as nuisance parameters such as angular spreading. Therefore the resulting algorithm involves a multi-dimensional search procedure over all unknown parameters.

In this paper a new method for DOA estimation of scattered sources is presented. The proposed method decomposes the channel vector into two orthogonal subspaces: the robust and non-robust subspaces. The robust subspace contains the part of the channel vector which is insensitive to the nuisance parameters. The estimator consists of two main stages: The first is identification of the robust and non-robust subspaces of the channel vector. In the second stage, this decomposition is used to define a new data model on which the ML estimator is developed. The resulting estimator involves a search only on the parameters of interest while it nulls the subspace which is sensitive to errors in the nuisance parameters.

## 2. PROBLEM FORMULATION

Consider an array of  $N$  sensors, monitoring a wave-field generated by a spatially distributed narrowband source in additive background noise. The complex envelope representation of the array output observation vector at the discrete time  $t_k$  can thus be modeled as

$$\mathbf{y}(t_k) = \mathbf{b}(\theta, t_k, \psi)s(t_k) + \mathbf{n}(t_k), \quad k = 1, \dots, K, \quad (1)$$

where  $\mathbf{b}(\theta, t_k, \psi)$  is the channel vector formed between the source and the array elements,  $s(t_k)$  is the transmitted signal and  $\mathbf{n}(t_k)$  is the additive noise vector.  $K$  is the number of available independent snapshots. The source signal and the additive noise are assumed to be independent, zero-mean, Gaussian random processes. The noise is modeled as spatially white:

$$E\{\mathbf{n}(t_k)\mathbf{n}^H(t_k)\} = \sigma_n^2 \mathbf{I}.$$

The unknown parameter,  $\theta$  is the mean direction and  $\psi$  represents the spatial source parameters, such as the source spreading. In most problems, the mean direction,  $\theta$ , is the parameter of interest while  $\psi$  consists of the unknown nuisance parameters. For a distributed source, the channel vector is modeled as a random vector such that:

$$\mathbf{b}(\theta, t_k, \psi) = \int_{-\pi}^{\pi} f(\phi, t_k | \theta, \psi) \mathbf{a}(\phi) d\phi \quad (2)$$

where  $f(\phi, t_k | \theta, \psi)$  is a complex random spatio-temporal weighting function which represents the local scattering. This model was proposed and used in [1], [4]. However, under this model, the estimator involves a multi-dimensional search over the vector of unknown parameters,  $\psi$ .

Our goal here is to estimate the mean direction of the source,  $\theta$ , from the data  $\{\mathbf{y}(t_k)\}_{k=1}^K$  in the presence of unknown nuisance parameters, such as the spatial source parameters,  $\psi$ , and the signal variance.

### 3. THE PROPOSED ESTIMATOR

In this section, an estimator of the nominal source location which is robust to spatial source distribution, is presented. It consists of two main stages: The first is identification of the robust and non-robust subspaces of the data. In the second stage, this decomposition is used to derive a model for which the ML estimator is developed. The resulting estimator involves a search procedure over the parameter of interest only, while it projects the received data into the subspace which is sensitive to uncertainties in the nuisance parameters.

#### 3.1. Identification of the robust and non-robust subspaces

Assuming that the nuisance vector parameter,  $\psi$ , is unknown random, the channel vector  $\mathbf{b}(\theta, t_k, \psi)$  can be decomposed as follows:

$$\mathbf{b}(\theta, t_k, \psi) = \bar{\mathbf{b}}(\theta) + \Delta\mathbf{b}(\theta, t_k, \psi) \quad (3)$$

where  $\bar{\mathbf{b}}(\theta)$  denotes the mean of  $\mathbf{b}(\theta, t_k, \psi)$  with respect to the random parameter,  $\psi$ :  $\bar{\mathbf{b}}(\theta) = E_{\psi}(\mathbf{b}(\theta, t_k, \psi))$ , where  $E_{\psi}$  denotes expectation with respect to the random parameters  $\psi$ . The term,  $\Delta\mathbf{b}(\theta, t_k, \psi)$  expresses

deviation of the channel vector from its average. By this notation, we assumed that the time-varying channel is a stationary process which may be a result of channel fluctuations, and therefore the statistics of the channel does not depend on  $t_k$ .

The average channel vector  $\bar{\mathbf{b}}(\theta)$  can be evaluated off-line using Monte-Carlo method for a given grid of  $\theta$  according to the statistics of  $\psi$ .

Let  $\mathbf{C}_b(\theta)$  denote the covariance of  $\Delta\mathbf{b}(\theta, t_k, \psi)$ :

$$\mathbf{C}_b(\theta) = E_{\psi}(\Delta\mathbf{b}(\theta, t_k, \psi)\Delta\mathbf{b}^H(\theta, t_k, \psi)).$$

Decomposition of the robust and non-robust subspaces is performed according to the covariance matrix  $\mathbf{C}_b(\theta)$ . Given sufficiently large number of array elements and small spread of the source, the matrix  $\mathbf{C}_b(\theta)$  is low rank:  $\text{rank}(\mathbf{C}_b(\theta)) = N_o < N$ , where  $N_o$  can be determined by a rank test of the covariance matrix. Thus, the matrix of its eigenvectors,  $\mathbf{H}(\theta)$ , can be decomposed as:

$$\mathbf{H}(\theta) = [\mathbf{H}_1(\theta) \quad \mathbf{H}_2(\theta)] \quad (4)$$

where the  $N \times N_o$  matrix,  $\mathbf{H}_1(\theta)$ , denotes its principal subspace. Therefore the channel vector deviation,  $\Delta\mathbf{b}(\theta, \psi)$ , can be approximated by the following representation:

$$\Delta\mathbf{b}(\theta, t_k, \psi) \approx \mathbf{H}_1(\theta)\beta(\theta, t_k, \psi) \quad (5)$$

The matrix  $\mathbf{H}_1(\theta)$  represents the most sensitive subspace of the channel vector to the source spreading. Now, the channel vector can be represented as

$$\begin{aligned} \mathbf{b}(\theta, t_k, \psi) &\approx \bar{\mathbf{b}}(\theta) + \mathbf{H}_1(\theta)\beta(\theta, t_k, \psi) \\ &= \underbrace{[\bar{\mathbf{b}}(\theta) \quad \mathbf{H}_1(\theta)]}_{\mathbf{B}(\theta)} \underbrace{\begin{bmatrix} 1 \\ \beta(\theta, t_k, \psi) \end{bmatrix}}_{\gamma(\theta, t_k, \psi)} \end{aligned} \quad (6)$$

In order to obtain estimators which avoid a search procedure over  $\psi$ , in the following the dependence of  $\gamma(\cdot, \cdot, \cdot)$  on  $\psi$  and  $\theta$  is ignored and  $\gamma(\theta, t_k, \psi)$  is assumed to be an unknown vector. The importance of the above decomposition is that it isolates a subspace which does not depend on the nuisance parameters,  $\psi$ .

#### 3.2. The Maximum-Likelihood estimator

By substitution of the channel vector model from (6) into (1) the data model becomes

$$\mathbf{y}(t_k) = \mathbf{B}(\theta)\gamma(t_k)s(t_k) + \mathbf{n}(t_k) = \mathbf{B}(\theta)\mathbf{g}(t_k) + \mathbf{n}(t_k), \quad (7)$$

where  $\mathbf{g}(t_k) \triangleq \gamma(t_k)s(t_k)$ .

For the case of coherently distributed sources, with no channel fluctuations,  $\gamma(t_k)$  is time-independent, and therefore  $\mathbf{g}(t_k)$  is a zero-mean random vector whose covariance matrix is of rank one:

$$\mathbf{R}_g \triangleq E\{\mathbf{g}(t_k)\mathbf{g}^H(t_k)\} = \sigma_g^2 \mathbf{v}\mathbf{v}^H, \quad (8)$$

where  $\mathbf{v}$  is an unknown, deterministic vector. Under the assumption of independent, zero-mean, Gaussian signal and noise, the conditional probability density function (pdf) of the data is zero-mean, Gaussian, with covariance:

$$\mathbf{R}_y = E\{\mathbf{y}(t_k)\mathbf{y}^H(t_k)\} = \mathbf{B}(\theta)\mathbf{R}_g\mathbf{B}^H(\theta) + \sigma_n^2\mathbf{I}. \quad (9)$$

Substituting (8) into (9) gives:

$$\mathbf{R}_y = \sigma_n^2 (\mathbf{B}(\theta)\mathbf{v}\mathbf{v}^H\mathbf{B}^H(\theta)snr + \mathbf{I}). \quad (10)$$

where the signal-to-noise ratio is defined as  $snr \triangleq \frac{\sigma_g^2}{\sigma_n^2}$ . Now, the ML estimator of  $\theta$  can be written as

$$\hat{\theta}_{ML} = \arg \max_{\theta} \max_{\mathbf{v}, snr, \sigma_n^2} f(\mathbf{y}(t_1), \dots, \mathbf{y}(t_K) | \mathbf{v}, snr, \sigma_n^2) \quad (11)$$

where  $\mathbf{v}, snr, \sigma_n^2$  are nuisance parameters and the function  $f(\mathbf{y}(t_1), \dots, \mathbf{y}(t_K) | \mathbf{v}, snr, \sigma_n^2)$  is the joint pdf of the data given the unknown parameters. In the Appendix, it is shown that by taking the logarithm of (reflikelihood), and maximizing over the nuisance parameters, the ML estimator of  $\theta$  becomes:

$$\hat{\theta}_{ML} = \arg \max_{\theta} \max_i (\lambda_i(\theta) - \log \lambda_i(\theta)) \quad (12)$$

where  $\lambda_i(\theta)$  are the eigenvalues of the matrix  $\mathbf{B}^H(\theta)\mathbf{S}\mathbf{B}(\theta)$  and  $\mathbf{S}$  is the sample covariance matrix:

$$\mathbf{S} \triangleq \frac{1}{K} \sum_{k=1}^K \mathbf{y}(t_k)\mathbf{y}^H(t_k).$$

This estimator ignores some prior statistical information on the variance of the elements of  $\beta$  that may be available from the first stage: note that  $cov(\beta(\theta, t_k, \psi))$  is the matrix of eigenvalues of  $\mathbf{C}_b(\theta)$ . With the modified model of (7), the dependence on the uncertainties in the nuisance parameters are expressed linearly. Therefore, these uncertainties can be considered as an additive noise on which some prior statistical information may be available. However, this additive noise is not necessarily Gaussian. By assumption of Gaussianity, the proposed estimator can be extended to maximum *a-posteriori* probability estimator which considers this prior statistical information.

#### 4. SIMULATION RESULTS

To illustrate the results, consider a distributed source with Gaussian shape spreading with angular spread of  $\Delta$ . Assume an equally spaced 15 sensor linear array of inter sensor separation of  $\lambda/2$  where  $\lambda$  is the wavelength of the transmitting source. Further, consider the case where the mean DOA is  $30^\circ$  and  $\sigma_n^2 = 1$ .

Fig. 1 depicts the eigenvalues of the matrix  $\mathbf{C}_b(\theta)$  for different values of  $\theta$  where the spreading parameter

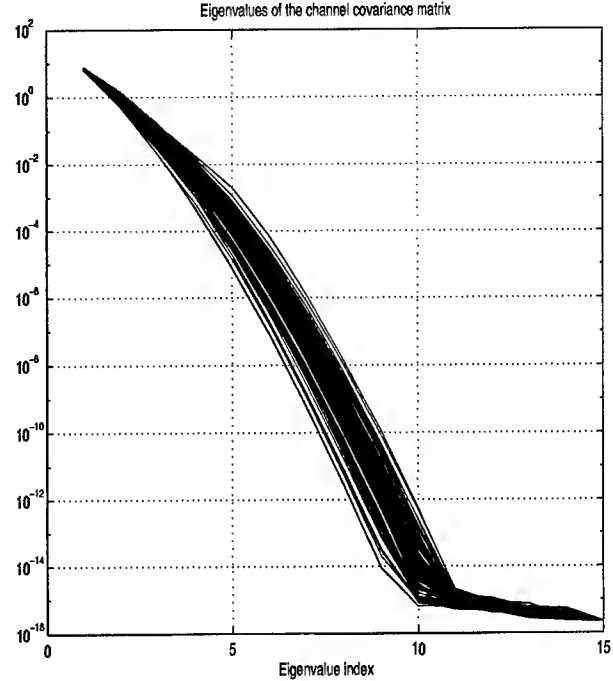


Figure 1: Eigenvalues of the channel vector covariance matrix,  $\mathbf{C}_b(\theta)$  for different DOA's,  $\theta$ .

was set to  $3^\circ$ . It shows that in this case, the channel covariance matrix is low rank, that is the source spreading causes perturbations of the channel vector in a limited subspace. Clearly, this subspace is larger for greater spreading parameter. This figure shows that in typical cases of a distributed source, there exists a large subspace which is insensitive to source spreading. This subspace enables source localization and it does not require estimation of the spreading parameters or spreading shape.

Fig. 2 demonstrates the performance of the robust ML method as a function of SNR for different source spreading parameter. In this example the number of snapshots is  $K = 100$ . As it was shown in [4], the estimation error does not converge to zero as the SNR goes to infinity.

Finally, Fig. 3 shows the performance of the proposed estimator as a function of the spreading parameter at SNR's above the threshold SNR. As expected, the error STD is an increasing function of the spreading parameter.

#### 5. CONCLUSIONS

In this paper, a new algorithm for scattered source localization is presented. The algorithm is robust to source spreading parameters and therefore does not require jointly estimating those nuisance parameters and the DOA which is usually the parameter of interest.

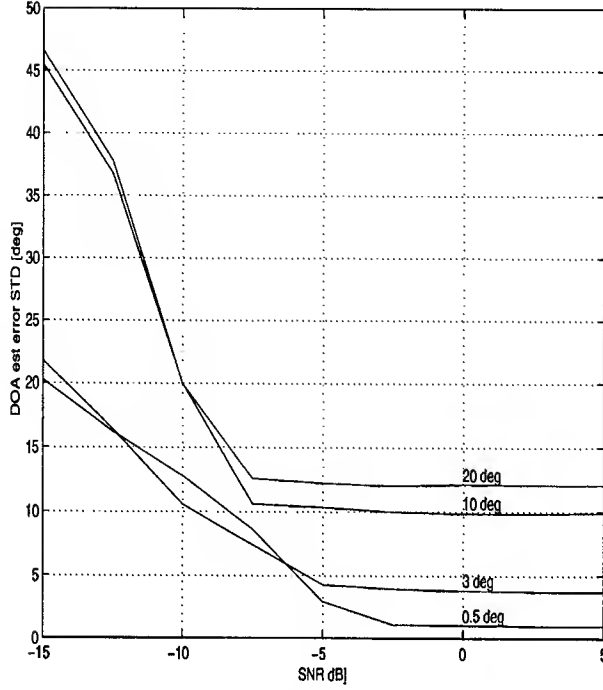


Figure 2: Performance of the proposed estimator as a function of SNR for different spreading levels.

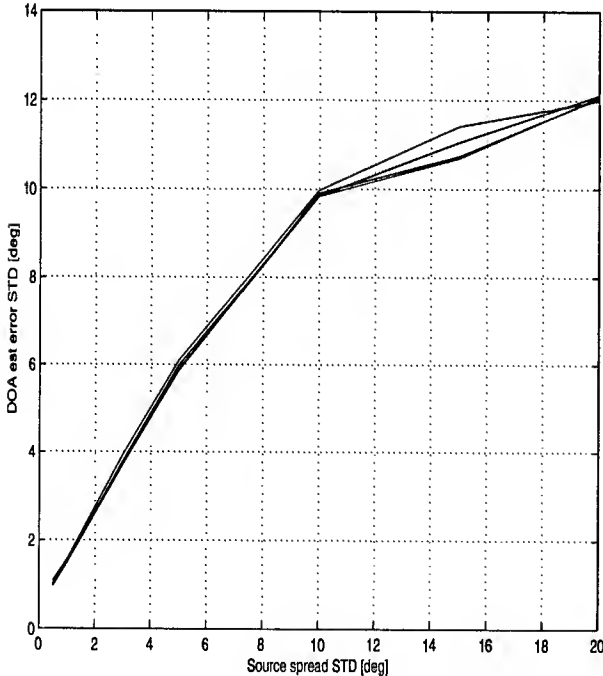


Figure 3: Performance of the proposed estimator as a function of the spreading parameter.

Therefore, it provides a computationally efficient technique for scattered source localization which does not involve any search over the spreading parameters. The proposed method is based on a decomposition of the channel vector to two subspaces and parameter estimation according to the subspace which is not sensitive to the source spreading parameters. An ML estimator for the decomposed model is presented and its performance is evaluated by Monte-Carlo simulations.

## A. APPENDIX

Derivation of the ML estimator: Eq. (12).

Taking the logarithm of the conditional pdf of (11), the ML estimator of  $\theta$  is

$$\hat{\theta}_{ML} = \arg \max_{\theta} \max_{\mathbf{v}, snr, \sigma_n^2} \left( -K \log \det(\mathbf{R}_y) - \sum_{k=1}^K \mathbf{y}^H(t_k) \mathbf{R}_y^{-1} \mathbf{y}(t_k) \right) \quad (\text{A.1})$$

From (10) it can easily verified that

$$\det(\mathbf{R}_y) = \sigma_n^{2N} (1 + snr \mathbf{v}^H \mathbf{B}^H(\theta) \mathbf{B}(\theta) \mathbf{v}) \quad (\text{A.2})$$

and

$$\mathbf{R}_y^{-1} = \frac{1}{\sigma_n^2} \left( \mathbf{I} - \frac{snr \mathbf{B}(\theta) \mathbf{v} \mathbf{v}^H \mathbf{B}^H(\theta)}{1 + snr \mathbf{v}^H \mathbf{B}^H(\theta) \mathbf{B}(\theta) \mathbf{v}} \right). \quad (\text{A.3})$$

By substituting (A.2) and (A.3) into (A.1), one obtains:

$$\hat{\theta}_{ML} = \arg \max_{\theta} \max_{\mathbf{v}, snr, \sigma_n^2} \left( -\log(1 + snr \mathbf{v}^H \mathbf{B}^H(\theta) \mathbf{B}(\theta) \mathbf{v}) + \frac{1}{\sigma_n^2} \frac{snr \mathbf{v}^H \mathbf{B}^H(\theta) \mathbf{S} \mathbf{B}(\theta) \mathbf{v}}{1 + snr \mathbf{v}^H \mathbf{B}^H(\theta) \mathbf{B}(\theta) \mathbf{v}} \right) \quad (\text{A.4})$$

where  $\mathbf{S}$  is the sample covariance matrix. Maximizing (A.4) with respect to  $snr$  gives

$$\hat{\theta}_{ML} = \arg \max_{\theta} \max_{\mathbf{v}} (G(\theta, \mathbf{v}) - \log(G(\theta, \mathbf{v}))) \quad (\text{A.5})$$

where  $G(\theta, \mathbf{v}) \triangleq \frac{\mathbf{v}^H \mathbf{B}^H(\theta) \mathbf{S} \mathbf{B}(\theta) \mathbf{v}}{\mathbf{v}^H \mathbf{B}^H(\theta) \mathbf{B}(\theta) \mathbf{v}}$ . With no loss of generality we assume that  $\|\mathbf{v}\|_{\mathbf{B}^H(\theta) \mathbf{B}(\theta)}^2 = \mathbf{v}^H \mathbf{B}^H(\theta) \mathbf{B}(\theta) \mathbf{v} = 1$ . Now, it is required to solve the following problem

$$\mathbf{v}^H \mathbf{B}^H(\theta) \mathbf{S} \mathbf{B}(\theta) \mathbf{v} \longrightarrow \max$$

with respect to  $\mathbf{v}$ , subject to the constraint:

$$\mathbf{v}^H \mathbf{B}^H(\theta) \mathbf{B}(\theta) \mathbf{v} = 1.$$

This maximization can be performed using Lagrange multipliers. Finally, the ML estimator can be written as:

$$\hat{\theta}_{ML} = \arg \max_{\theta} \max_i (\lambda_i(\theta) - \log \lambda_i(\theta)) \quad (\text{A.6})$$

where  $\lambda_i(\theta)$  are the eigenvalues of the matrix  $\mathbf{B}^H(\theta) \mathbf{S} \mathbf{B}(\theta)$ .

## REFERENCES

- [1] S. Valaee, B. Champagne and P. Kabal, "Parametric localization of distributed sources," *IEEE Trans. on SP*, vol. 43, no. 9, pp. 2144-2153, September 1995.
- [2] D. Astely, B. Ottersten and A. L. Swindlehurst, "Generalised array manifold model for wireless communication channels with local scattering," *IEE Proc. Radar, Sonar Navig.*, vol. 145, no. 1, pp. 51-57, February 1998.
- [3] D. Astely and B. Ottersten, "The effect of local scattering in direction of arrival estimation with MUSIC," *IEEE Trans. on SP*, vol. 47, no. 12, pp. 3220-4234, December 1999.
- [4] R. Raich, J. Goldberg and H. Messer, "Localization of a distributed source which is "partially coherent" - modeling and Cramer-Rao bounds," *Proc. of ICASSP99*, March 1999.
- [5] T. Abdellatif, P. Larzabal and H. Clergeot, "Performance study of a generalized subspace-based method for scattered sources," *Proc. of ICASSP2000*, June 2000.

# ISAR IMAGING AND CRYSTAL STRUCTURE DETERMINATION FROM EXAFS DATA USING A SUPER-RESOLUTION FAST FOURIER TRANSFORM

*George Zweig*

Signition, Inc.  
190 Central Park Square  
Los Alamos, NM 87544  
zweig@signition.com

*Brendt Wohlberg*

Theory Division and  
Center for Nonlinear Studies  
Los Alamos National Laboratory  
Los Alamos, NM 87545

## ABSTRACT

The method of sparse spectrum estimation developed by Chen and Donoho for real-valued one-dimensional signals [2] has been extended to complex-valued signals [12], and is used here in two widely different applications: to denoise and superresolve ISAR images, and to transform extended X-ray absorption fine-structure (EXAFS) data of the elements to aid in the determination of their detailed crystal structure. This extension of the Chen-Donoho algorithm, which we call the  $l^1$ -FFT, incorporates the a priori information that the spectrum is sparse by minimizing the  $l^1$  norm of the coefficients of the expansion functions. The  $l^1$ -FFT is applied to stepped-frequency ISAR imaging where it increases resolution by factors of 4 and 64 over that of the windowed Fourier transform, for the real and simulated data presented here. In the second application, to determine the effects of aging on the crystal structure of plutonium, the  $l^1$ -FFT is used to transform EXAFS plutonium data. The  $l^1$ -FFT increases inter-atomic spatial resolution by a factor of 64 over that delivered by a windowed Fourier transform.

## 1. INTRODUCTION TO THE $l^1$ -FFT

Recently Chen and Donoho presented a novel method of super-resolution spectrum estimation for real-valued one-dimensional signals [2]. They write the signal as an overcomplete linear combination of sinusoids with many more frequencies than points in the signal. To limit the number of possible expansions, they choose the coefficients of the sinusoids in the expansion so that the sum of their absolute values is as small as possible, subject to the constraint that the sum of sinusoids adds up to the signal at its sampled values. Technically, the  $l^1$  norm of the expansion is minimized. Minimizing the  $l^1$  norm favors expansions with fewer large terms over many small terms. The Method of Frames, which minimizes the  $l^2$  norm, does just the opposite [5]. Even if the function being expanded is just a single expansion function, every other expansion function generally has a nonzero inner product with it, i.e., a nonzero coefficient in its Method of Frames expansion.

Minimizing the  $l^p$  length for any  $p$  in the range  $0 < p < 1$  favors a sparse representation of the signal, but if the  $p = 1$  norm is chosen, a global minimum for the expansion coefficients can be found efficiently using re-

cently developed fast linear programming algorithms [3]. Because linear programming optimizes globally, it can stably superresolve in ways that Matching Pursuits [7, 9] cannot.

The  $l^0$  case, with its local minima and bounds on the deviations from the global minimum, is discussed by Natarajan [8]. Other methods for obtaining sparse representations are described in references [6] and [10].

*Denoising by relaxing constraints:* If the signal has noise added to it, the resulting noisy signal should not be represented exactly as a sparse sum of sinusoids. In the simplest method for denoising, the requirement that the weighted sinusoids sum exactly to the noisy signal is relaxed [1]. The deviation allowed is set by the signal-to-noise ratio.

*Denoising by including delta functions as expansion functions:* In a second method of denoising, delta functions situated at every sampled point are added to the expansion set of sinusoids. The overall  $l^1$  norm of the expansion containing both sinusoids and delta functions is minimized. The signal is estimated by summing only the sinusoids, the noise by summing only the delta functions [1]. The relative amplitudes of the sinusoids and delta functions in the expansion set depends on the signal-to-noise ratio.

In the next sections the Chen-Donoho algorithm for signals with nonnegative coefficients is outlined, the  $l^1$ -FFT for real and complex signals is referenced, and the  $l^1$ -FFT is applied to ISAR imaging and the determination of crystal structure from EXAFS data.

### 1.1. Sparse representations with $l^1$ -norm minimization

In order to describe the Chen-Donoho method of sparse representation, consider the problem of expanding a sampled signal  $x[k]$  into a linear combination of expansion functions  $w_n[k]$ , with

nonnegative expansion coefficients  $\hat{x}[n]$ ,

$$x[k] = \sum_{n=0}^{N-1} \hat{x}[n]w_n[k], \quad k = 0, 1, \dots, K-1. \quad (1)$$

Assume that there are more expansion functions than points in the signal, i.e.,  $N > K$ . Then the expansion for  $\mathbf{x}$  is not unique and the coefficients  $\hat{x}[n]$  are underdetermined. Fix them by minimizing their  $l^1$  norm, i.e.,

$$\text{minimize } \|\hat{\mathbf{x}}\|_1 \equiv \sum_{n=0}^{N-1} |\hat{x}[n]|. \quad (2)$$

This leads to a sparse representation of  $\mathbf{x}$ , i.e., the fraction of coefficients  $\hat{x}[n]$  that are large will be small relative to the fraction that are large if the  $l^2$  norm were minimized (the Method of Frames [5]).

In matrix form, find a vector  $\hat{\mathbf{x}}$  that will

$$\text{minimize } \|\hat{\mathbf{x}}\|_1, \quad \text{subject to } \mathbf{W}\hat{\mathbf{x}} = \mathbf{x}, \quad (3)$$

where

$$\hat{\mathbf{x}} \geq 0. \quad (4)$$

The matrix element  $w_{kn}$  is the  $k$ th sample of the  $n$ th expansion function. This is a linear programming problem. A method for solving Eq. 3 when  $\hat{\mathbf{x}}$  can be either positive or negative is given in reference [1], and the complex coefficient case is described in [12].

## 2. ISAR IMAGING

The application of the  $l^1$ -FFT to ISAR imaging is given in Figures 1 and 2, where simulated Mig 25 data and real Boeing 727 data are used to form superresolution ISAR images.

## 3. EXAFS CRYSTAL STRUCTURE

An understanding of how the crystal structure of plutonium changes while aging may be helpful in establishing the functionality and safety

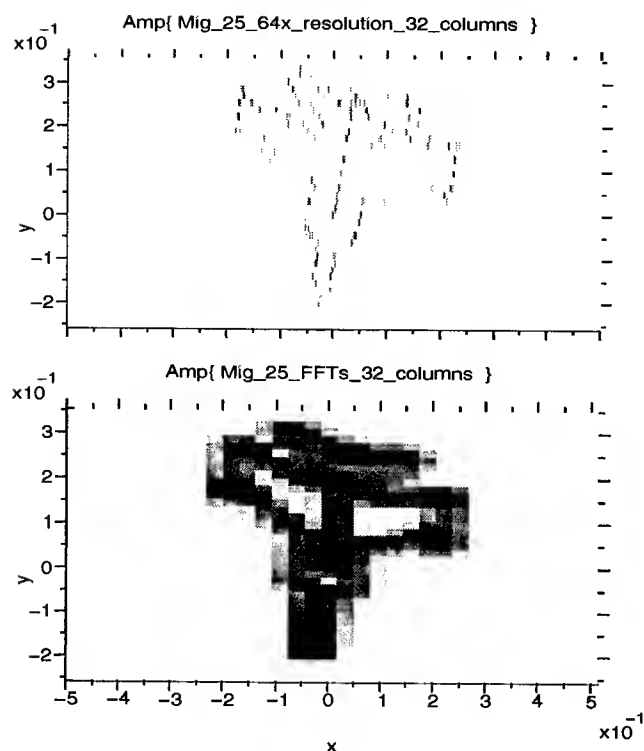


Figure 1: *Top*: ISAR image from simulated Mig 25 data using the  $l^1$ -FFT. Each row of the image was constructed from 32 data points. A total of 512 points were available, but since the plane was accelerating only 1/16th of the data was used to create an “instantaneous” snapshot. The resolution is 64 times greater than that provided by a Fourier transform. No delta functions were used in the fit because no noise was added in the simulation. Hann-windowed Fourier transforms were first used to transform the data column-wise. *Bottom*: The same data used to create the top image was analyzed row-wise with Hann-windowed Fourier transforms. Note the difference of a factor of 64 in the horizontal resolution. The vertical resolution in the top and bottom images is identical because the same column-wise processing was used for both images.

Amp{ 727\_real\_data\_4x\_resolution\_32\_columns }

Amp{ 727\_Hann\_FFTs\_32\_columns }

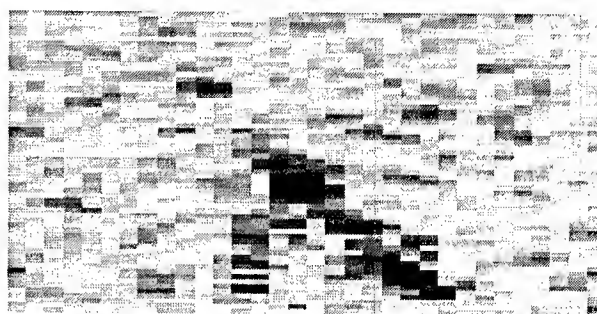


Figure 2: *Top*: ISAR image from real 727 data using the  $l^1$ -FFT. Each row of the image was constructed from 32 data points. The resolution is 4 times greater than that provided by a Fourier transform. Greater resolution was obtained in other ISAR images formed from the same data, but the increased resolution created “inferior” images for viewing because the resulting line segments from which the images were constructed were too thin to have significant visual impact. Delta functions were used in the fit to remove essentially all noise. Hann-windowed Fourier transforms were first used to transform the data column-wise. *Bottom*: The same data used to create the top image was analyzed row-wise with Hann-windowed Fourier transforms. Note the difference of a factor of 4 in the horizontal resolution and the great reduction in noise in the top image that results from including delta functions in the expansion set. The vertical resolution in the top and bottom images is identical because the same column-wise processing was used for both images.



of stockpiled nuclear weapons. The fine structure observed in the energy dependence of the total absorption cross section of X-rays in plutonium can be used to determine the relative distances between neighboring atoms in plutonium, since the fine structure reflects the interference of electrons directly ionized by the X-rays with those undergoing additional scattering off neighboring atoms after initial ionization [11].

The X-ray absorption interference term  $I(k)$  at photon momentum  $k$  is a sum over all paths taken by the ionized electron. In the single-scattering approximation, when the ionized electron is approximated by a plane wave,  $I(k)$  becomes

$$I(k) \equiv \sum_{\text{sites } s} \frac{A_s(k)}{kr_s^2} e^{-2(r_s/\lambda(k) + \sigma_s^2 k^2)} e^{i[2kr_s + \phi_s(k)]}, \quad (5)$$

where  $A_s(k)$  and  $\phi_s(k)$  are the scattering amplitude and phase of the outgoing electron scattering off the  $s$ th site at distance  $r_s$  from the point of ionization,  $\lambda(k)$  is the electron mean free path, and  $\sigma_s^2$  is the Debye-Waller factor resulting from phonon motion. The magnitude of the electron mean free path  $\lambda(k)$  restricts the number of important scattering sites to those in the neighborhood of the ionized atom.

Theory provides estimates of all quantities except the inter-site distances [11], which can ultimately be found with the  $l^1$ -FFT [13].

Figure 3 compares the windowed FFT amplitude and the  $l^1$ -FFT (LIFT) amplitude of newly-processed plutonium EXAFS data. The top panel shows the same two curves as the bottom panel, but with an expanded ordinate. The spatial resolution provided by the  $l^1$ -FFT is 64 times higher than that of the windowed FFT. All peaks in the  $l^1$ -FFT amplitude, except for the peak at 3.9 Å, correspond to the positions of atoms in  $\delta$ -plutonium (face-centered cubic lattice). Evidence for an extra (anomalous) site in newly-processed plutonium has been

reported by Conradson using a windowed FFT [4]. His extra peak was located at 3.7 Å, but with poorer spatial resolution. Although the difference between 3.9 and 3.7 Å in site-position is numerically small, theoretical models that try to explain the existence of this extra peak are very sensitive to its exact location. Therefore an accurate experimental determination of this site-position is important.

#### 4. ACKNOWLEDGEMENTS

Dr. Victor Chen has kindly provided the ISAR data used in this report. We also thank Steve Conradson for providing the plutonium EXAFS data, and Richard Silver for putting it into a form appropriate for Fourier analysis.

#### REFERENCES

- [1] S. S. Chen. *Basis Pursuit*. PhD thesis, Stanford Univ., Stanford, CA, 1995.
- [2] S. S. Chen and D. L. Donoho. Application of basis pursuit in spectrum estimation. In *IEEE Int. Conf. Acoust., Speech, Signal Proc.* IEEE Service Center, Piscataway, 1998.
- [3] S. S. Chen, D. L. Donoho, and M. A. Saunders. Atomic decomposition by basis pursuit. *SIAM J. Sci. Comput.*, 20:33–61, 1998.
- [4] S. Conradson. Application of XAFS to materials and environmental science. *Applied Spectroscopy*, 52:252A–279A, 1998.
- [5] I. Daubechies. *Ten Lectures on Wavelets*. Society for Industrial and Applied Mathematics, Philadelphia, 1992.
- [6] G. Harikumar and Y. Bresler. A new algorithm for computing sparse solutions to linear inverse problems. In *IEEE Int.*

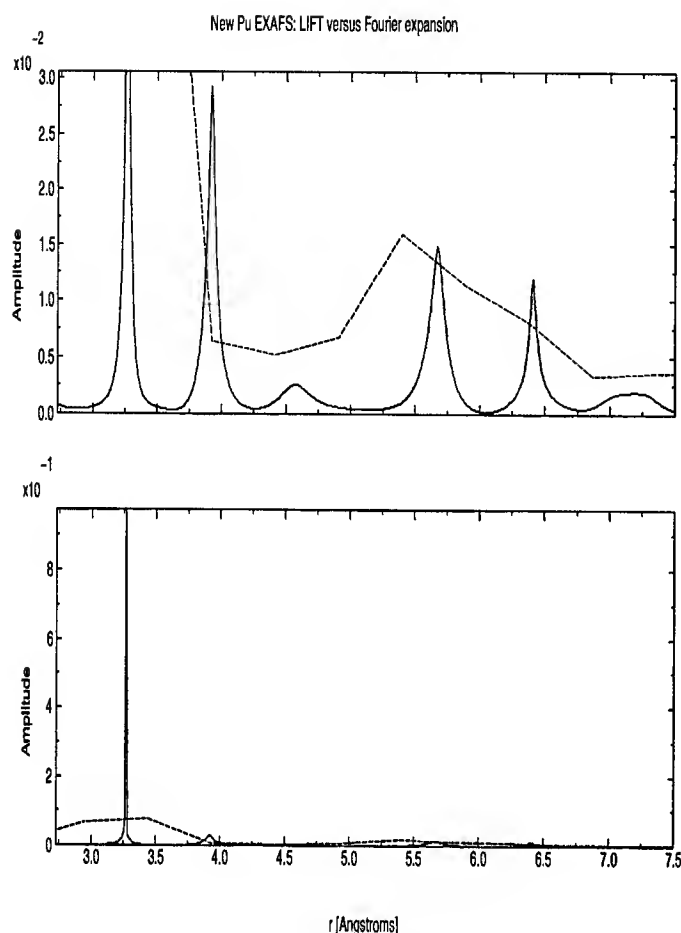


Figure 3: The windowed FFT amplitude (dashed line) and the  $l^1$ -FFT amplitude (solid line, referred to as LIFT in the Figure title) of newly-processed plutonium EXAFS data. The top panel expands the ordinate of the lower panel. The peaks correspond to the positions of plutonium atoms. A face-centered cubic lattice structure predicts the positions of all the peaks observed, except for the peak at 3.9 Å. The  $l^1$ -FFT increases spatial resolution by a factor of 64.

*Conf. Acoust., Speech, Signal Proc.*, volume III, pages 1331–1334. IEEE Service Center, Piscataway, 1996.

- [7] S. Mallat and Z. Zhang. Matching pursuits with time-frequency dictionaries. *IEEE Trans. on Signal Processing*, 41:3397–3415, 1993.
- [8] B. K. Natarajan. Sparse approximate solutions to linear systems. *SIAM J. Comput.*, 24:227–234, 1995.
- [9] S. Qian and D. Chen. Signal representation via normalized gaussian functions. *Signal Processing*, 36:1–11, 1994.
- [10] B. D. Rao and K. Kreutz-Delgado. An affine scaling methodology for best basis selection. *IEEE Trans. on Signal Proc.*, 47:187–200, 1999.
- [11] J. J. Rehr and R. C. Albers. Theoretical approaches to X-ray absorption fine structure. *Rev. Mod. Phys.*, 2000. Accepted for publication.
- [12] G. Zweig. Multi-window spectral estimation with optical processing for ISAR imaging, April 1999. 1st ONR SBIR Phase II progress report, Contract N00014-98-C-0196.
- [13] G. Zweig and B. Wohlberg. Age-related changes in Pu deduced from EXAFS data with a new super-resolution fast Fourier transform. In *Proceedings of the Nuclear Explosives Design Physics Conference*, volume XII. Los Alamos National Laboratory, Los Alamos, N.M., October 1999.

# Analysis of Radar Micro-Doppler Signature With Time-Frequency Transform

Victor C. Chen

Naval Research Laboratory, Washington DC 20375

## ABSTRACT

Micro-Doppler induced by mechanical vibrating or rotating of structures in a radar target is potentially useful for target detection, classification and recognition. While the Doppler frequency induced by the target body is constant, the micro-Doppler due to vibrating or rotating structures of the target is a function of dwell time. Analysis of the time-varying Doppler signature in the joint time-frequency domain can provide useful information for target detection, classification and recognition.

## INTRODUCTION

Mechanical vibration or rotation of structures in a target may induce frequency modulation on returned signals and generate side-bands about the center frequency of the target's body Doppler frequency. The modulation due to vibrations, which is usually at very low frequencies relative to the body Doppler frequency, is called micro-Doppler phenomenon. The modulation induced by rotations, which can be seen as a special case of vibrations and may have higher frequencies relative to the body Doppler frequency, may also be called the micro-Doppler. The micro-Doppler phenomenon can be regarded as a signature of the interaction between the vibrating or rotating structures and the body of the target and provides an additional information for target recognition complementary to existing recognition methods.

In coherent radar, the phase of the returned signal from a target is sensitive to variation in range. A half wavelength's change can cause 360° phase change. It is conceivable that the vibration of a reflecting surface may be measured with the phase change. Thus, the Doppler frequency shift, that represents the change of phase function with time, can be used to detect vibrations or rotations of structures in a target.

Figure 1 illustrates a reflector illuminated by radar located at the origin of a  $(x, y, z)$  coordinate system. The reflector  $P$  is vibrating about a center point  $Q$  at a distance  $R_0$  from the radar. If the azimuth and elevation angle of the point  $Q$  relative to the radar is  $\alpha$  and  $\beta$ , respectively, the point  $Q$  is at  $(R_0 \cos \beta \cos \alpha, R_0 \cos \beta \sin \alpha, R_0 \sin \beta)$  in the  $(x, y, z)$  coordinates. Assume that the reflector is at a distance  $D_t$  from the point  $Q$  that is also the origin of a coordinates  $(x', y', z')$  translated from  $(x, y, z)$ . If the azimuth and elevation angles of the reflector  $P$  relative to the center point  $Q$  is  $\alpha_p$  and  $\beta_p$ , respectively, the reflector will be at  $(D_t \cos \beta_p \cos \alpha_p, D_t \cos \beta_p \sin \alpha_p, D_t \sin \beta_p)$  in the  $(x', y', z')$  coordinates. Therefore, the vector from the radar to the reflector

becomes  $\bar{r}_t = \bar{R}_0 + \bar{D}_t$  as shown in the Fig.1. Generally, the range from the radar to the reflector can be expressed as

$$r_t = |\bar{r}_t| = [(R_0 \cos \beta \cos \alpha + D_t \cos \beta_p \cos \alpha_p)^2 + (R_0 \cos \beta \sin \alpha + D_t \cos \beta_p \sin \alpha_p)^2 + (R_0 \sin \beta + D_t \sin \beta_p)^2]^{1/2}$$

In the case that the azimuth angle  $\alpha$  of the point  $Q$  and the elevation angle  $\beta_p$  of the reflector  $P$  are all zero, we have

$r_t = (R_0^2 + D_t^2 + 2R_0 D_t \cos \beta \cos \alpha_p)^{1/2} \approx R_0 + D_t \cos \beta \cos \alpha_p$  for  $R_0 \gg D_t$ . If the vibration rate of the reflector is  $\omega_v$  and the amplitude of the vibration is  $D_v$ , the range of the reflector becomes

$$r(t) = r_t = R_0 + D_v \sin \omega_v t \cos \beta \cos \alpha_p$$

Thus, the received radar signal becomes

$$s(t) = \rho \exp\{j[2\pi f_c t + 4\pi \frac{r(t)}{\lambda_c}]\} = \rho \exp\{j[2\pi f_c t + \phi(t)]\}$$

where  $\phi(t) = 4\pi r(t)/\lambda_c$  is the phase function.

Because the time-derivative of the phase is frequency, by taking the time-derivative of the phase, the micro-Doppler frequency induced by the vibration is

$$f_D = \frac{4\pi}{\lambda_c} D_v \omega_v \cos \beta \cos \alpha_p \cos \omega_v t$$

The maximum of the Doppler frequency change is  $(4\pi/\lambda_c) D_v \omega_v$ , that can be reached when the orientation of the vibrating reflector is along the projection of the radar line-of-sight direction, i.e.,  $\alpha_p = 0$ , and the elevation angle  $\beta$  of the reflector is also 0.

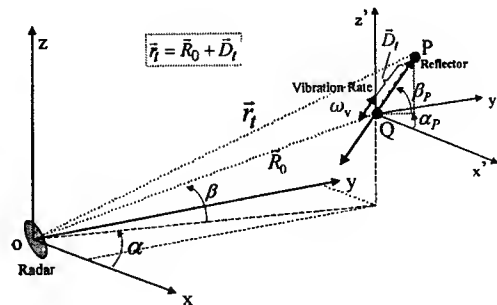
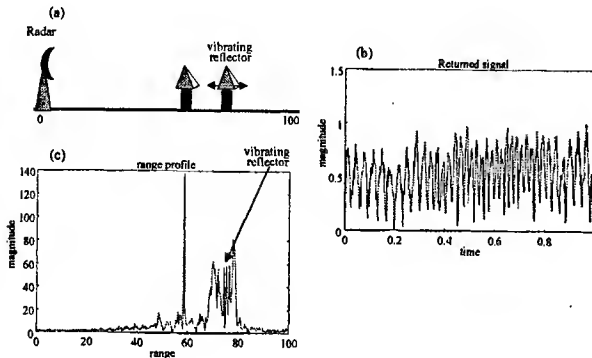


Figure 1. Geometry of a radar and a vibrating reflector.

# 1. FREQUENCY DOMAIN SIGNATURE

By taking the Fourier transform of the radar returned signal, the micro-Doppler frequency shift may be observed in the frequency domain [1]. Fig.2(a) illustrates a radar and two corner reflectors separated in a distance of 13.5m where one is stationary and the other is vibrating at 1.5Hz with a displacement of 3cm. Fig.2(b) is the returned I and Q signals from the two reflectors. The range profile is obtained by taking the Fourier transform of the returned I and Q signals that is shown in Fig.2(c). The spread peak in the range profile indicates that there may be a vibrating reflector in that region.

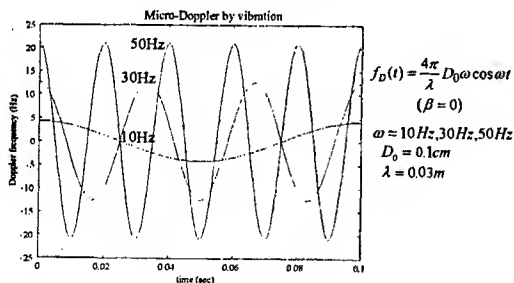


**Figure 2.** An experimental radar data (b) return from two corner reflectors (a): one is stationary and the other is vibrating. The Fourier transform of the returned signal is the range profile (c).

## 2. TIME-FREQUENCY DOMAIN SIGNATURE

### 2.1 VIBRATION-INDUCED MICRO-DOPPLER

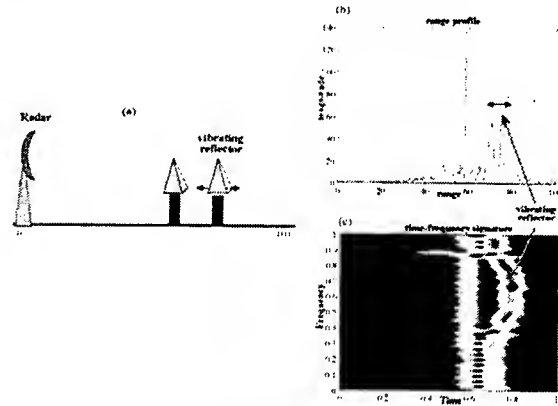
As shown in Fig.3, when radar is operating at X-band with 0.03m wavelength, a vibration at 10Hz with a displacement of 0.1cm will induce a maximal micro-Doppler frequency shift of 4.2Hz, which is detectable with a high-resolution radar.



**Figure 3.** Micro-Doppler generated by a vibrating reflector.

Fig.4 shows the time-frequency domain signature of the X-band experimental radar data returned from the two reflectors mentioned earlier. Fig.4(a) is the magnitude of the radar received I & Q data, and 4(b) shows the time-frequency signature of the data, where the vibration can be observed very

well. The time-frequency signature is obtained by taking time-frequency transforms, such as the Gabor transform [2]. From the time-frequency signature of the vibration we can estimate the vibration rate and, also, re-focus the vibrating reflector by taking time samples.

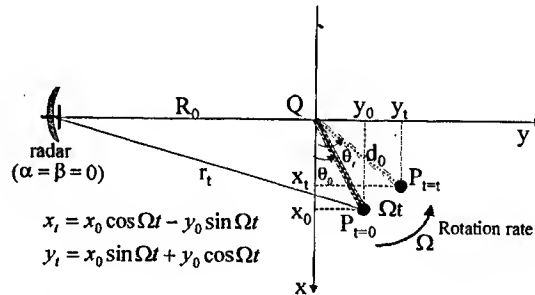


**Figure 4.** Time-frequency domain analysis of the returned signal.

### 2.2 ROTATION-INDUCED MICRO-DOPPLER

Rotating parts, such as helicopter's rotor blades and rotating antennas, are in rotational motion that will impart a periodic modulation on the returned signals from the rotating structures. The periodic modulation can generate a radar signature that can be used for target identification.

Helicopter's rotor blade can be modeled as a rigid, homogeneous linear antenna [3]. The electromagnetic backscattering signal from a point on the antenna has a Doppler frequency shift from the Doppler frequency of the rotor center.



**Figure 5.** Geometry of a radar and a rotating reflector.

Let us begin with a simple case as shown in Fig.5 where a scatterer  $P$  from one rotor blade rotates about a center point  $Q$  with a rotation rate of  $\Omega$ . The distance from the scatterer to the center point is  $d_0$ , and the distance between the radar and the center point is  $R_0$ . If both the radar and the rotor are on the same 2-D plane, i.e. the elevation angle is zero, the range from the radar to the scatterer becomes

$$r(t) \cong R_0 + vt + d_0 \sin \theta_0 \cos \Omega t + d_0 \cos \theta_0 \sin \Omega t$$

where  $v$  is the radial velocity of the helicopter and  $\theta_0$  is the initial rotation angle of the scatterer. The radar received signal becomes

$$s(t) = \rho \exp\{j[2\pi f_c t + \frac{4\pi}{\lambda_c} r(t)]\} = \rho \exp\{j[2\pi f_c t + \phi(t)]\}$$

where  $\phi(t) = 4\pi r(t)/\lambda_c$  is the phase function. In this case, the Doppler frequency shift of the scatterer can be obtained as

$$f_D = \frac{d}{dt}\phi(t) \cong \frac{4\pi}{\lambda_c} d_0 \Omega (-\sin \theta_0 \sin \Omega t + \cos \theta_0 \cos \Omega t)$$

If the rotor has an elevation angle  $\beta$ , then the above equation can be modified as

$$f_D \cong \frac{4\pi}{\lambda_c} d_0 \Omega \cos \beta (-\sin \theta_0 \sin \Omega t + \cos \theta_0 \cos \Omega t)$$

For N-blade rotor, assuming that one scatterer represents one blade, there are total N scatterers at different initial rotation angles:

$$\theta_k = \theta_0 + k2\pi/N, (k = 0, 1, 2, \dots, N-1)$$

and the total received signal becomes

$$\begin{aligned} s(t) &= \sum_{k=0}^{N-1} s_k(t) = \sum_{k=0}^{N-1} \rho_k \exp\{j[2\pi f_c t + \phi_k(t)]\} \\ &= \exp\{j2\pi f_c t\} \sum_{k=0}^{N-1} \rho_k \exp\{j\phi_k(t)\} \end{aligned}$$

where

$$\begin{aligned} \phi_k(t) &= \frac{4\pi}{\lambda_c} r_k(t) \\ &= \frac{4\pi}{\lambda_c} [R_0 + vt + \cos \beta (d_0 \sin \theta_k \cos \Omega t + d_0 \cos \theta_k \sin \Omega t)] \\ &= \frac{4\pi}{\lambda_c} [R_0 + vt + d_0 \cos \beta \sin(\Omega t + \theta_0 + k2\pi/N)] \quad (k = 0, 1, 2, \dots, N-1) \end{aligned}$$

By taking the Fourier transform, the frequency spectrum of the received signal can be expressed as

$$S(f) = C_0 \delta(f - f_c) + \sum_{k=0}^{N-1} C_k [\delta(f - f_c - kN\Omega) + \delta(f - f_c + kN\Omega)]$$

where  $C_0$  and  $C_k$  are determined by  $\lambda_c, R_0, v, d_0, \beta, N, \theta_0$ , and  $\Omega$  and may be defined as Bessel functions [4]. The first term is the carrier frequency and the terms in the summation determine the micro-Doppler generated from the rotor blades. Fig.6(a) demonstrates a radar returned signal from rotating rotor blades, and 6(b) shows the frequency spectrum of the returned signal where we can see the micro-Doppler frequencies generated from the rotor blades.

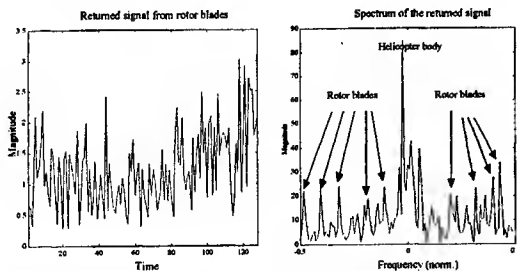


Figure 6. (a) Returned signal from rotor blades; (b) Spectrum of the returned signal.

There are many publications on the analysis of propeller modulation (PM) and jet engine modulation (JEM) in the frequency domain [3,5,6,7,8]. Fig.7 shows the time-frequency signature of the returned signal from rotor blades, where the characteristics of the rotating blades can be seen more clearly in the joint time-frequency domain. The strong time-frequency coefficients along the horizontal line about the center frequency are due to the returns from the helicopter's body. After we suppress the time-frequency coefficients of the helicopter body, the strong time-frequency coefficients along the dot-slope-lines are due to the returns from the rotating blades as shown in Fig.8. Because time information is available, the rotation rate of the blades can be measured from their time-frequency signature.

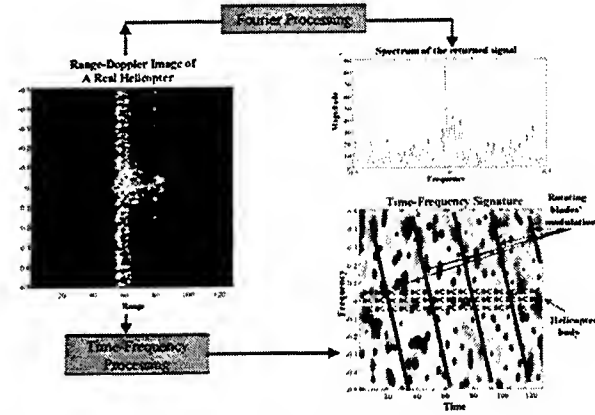


Figure 7. Time-frequency signature of the radar returned signal from rotating rotor blades.

## 2.3 WALKING MAN WITH SWINGING ARMS

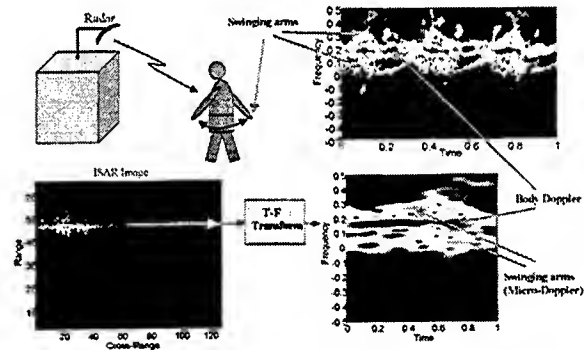


Figure 8. Micro-Doppler signature of a walking man with swinging arms.

Fig.8 illustrates a man walking towards a radar operating at X-band. ISAR image of the walking man is in range and cross-range domain. The hot spot is the body of the walking man. If we analyze the range profile at the range cell where the body is located in using a time-frequency transform, we can see the swinging arms. One arm has a Doppler frequency above the body's Doppler frequency and the other arm has a Doppler frequency below the body's Doppler. The upper-right picture

shows the superposition of the time-frequency signatures over several range profiles when the man is walking. We can see the body's Doppler frequency is almost constant and the arm's micro-Doppler becomes time-varying with a sinusoidal-like curve.

## 2.4 ROTATING ANTENNA

Fig.9 shows the micro-Doppler signature of a rotating antenna. The real part and the imaginary part of the radar return from the rotating antenna are shown in the figure. The time-frequency transform of the radar return is shown on the right where the time-frequency signature is unwrapped in the frequency domain. The parallel sloped lines are the micro-Doppler signature of the rotating antenna. From the time and frequency information, the rotation rate of the antenna can be calculated.

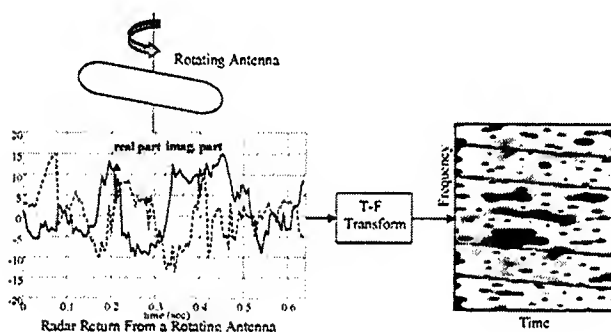


Figure 9. Rotating antenna and its micro-Doppler signature.

## 3. SUMMARY

We discussed the micro-Doppler phenomenon induced by mechanical vibrations or rotations of structures in a radar target, and proposed a time-frequency analysis of the micro-Doppler. The time-frequency signature of the micro-Doppler provides additional time information and shows micro-Doppler frequency variations with time. Thus, an additional information about vibration rate or rotation rate is available for target recognition.

## ACKNOWLEDGMENTS

This work was sponsored by the Office of Naval Research. We would also like to express our thanks to the DREO, Canada for the radar data of vibrating reflector, and the Norden Systems, Northrop Grumman for the radar data of walking man.

## REFERENCES

- [1] Wong, S.K., Kashyap, S., Louie, A., Gauthier, S., and Riseborough, E., "Target identification in the frequency domain", RTO Meeting Proceedings 6: Non-Cooperative Air Target Identification Using Radar, pp.19.1-19.14, April, 1998
- [2] Qian, S. and Chen, D., *Joint Time-Frequency Analysis- Methods and Applications*, Englewood Cliffs, NJ: Prentice-Hall, 1996
- [3] Schneider, H., "On the maximum entropy method for Doppler spectral analysis of the radar echoes from rotating objects", IEE International Radar Conference, pp.279-282, 1987
- [4] Willey, J. and Faust, H., "Spatial spectra for aircraft identification", RTO Meeting Proceedings 6: Non-Cooperative Air Target Identification Using Radar, pp.7.1-7.12, April, 1998
- [5] Bullard, B.D. and Dowdy, P.C., "Pulse Doppler signature of a rotary-wing aircraft", IEEE AES Systems Magazine, pp.28-30, May 1991
- [6] Fliss, G.G. and Mensa, D.L., "Instrumentation for RCS measurements of modulation spectra of aircraft blades", IEEE 1986 National Radar Conference, pp.95-99, 1986
- [7] Martin, J. and Mulgrew, B., "Analysis of the theoretical radar return signal from aircraft propeller blades", IEEE 1990 International Radar Conference, pp.569-572, 1990
- [8] Bell, M.R. and Grubbs, R.A., "JEM modeling and measurement for radar target identification", IEEE Trans. on AES, vol.29, no.1, pp.73-87, 1993

# ESTIMATING THE PARAMETERS OF MULTIPLE WIDEBAND CHIRP SIGNALS IN SENSOR ARRAYS

Alex B. Gershman\*

Marius Pesavento\*

Moeness G. Amin\*\*

\*Department of ECE, McMaster University  
Hamilton, L8S 4K1 Ontario, Canada  
gershman@ieee.org

\*\*Department of ECE, Villanova University,  
Villanova, PA 19085, USA

## ABSTRACT

The problem of estimating the parameters of multiple wideband polynomial-phase signal sources in sensor arrays is addressed. A new deterministic maximum likelihood (ML) direction of arrival (DOA) estimator and the respective Cramér-Rao bound (CRB) are presented for the general case of multiple constant-amplitude polynomial-phase sources. Since the proposed ML estimator is computationally intensive, an approximate solution is proposed, originating from the analysis of the ML function in the single chirp case. As a result, the so-called *chirp beamformer* is derived, which is applicable to "well-separated" sources that have distinct time-frequency or/and spatial signatures. Our beamforming approach requires solving a 3D optimization problem and, therefore, enjoys essentially simpler implementation than that dictated by the exact ML.

## 1. INTRODUCTION

Estimating the parameters of polynomial-phase signals is an important problem because linear FM (chirp) and nonlinear FM signals are encountered in many practical applications [1]-[3]. Recently, there has been a growing interest in estimating the parameters of multiple polynomial-phase signals in sensor arrays [4]-[7]. Several authors solved this problem using narrowband assumptions. In [5], a new spatial time-frequency distribution (STFD) concept has been developed and employed for direction finding of narrowband chirp sources using subspace techniques. Several exact and approximate ML algorithms for this estimation problem have been proposed [4]. Promising extensions of the above-mentioned narrowband approaches to the wideband polynomial-phase signal case have been recently reported [6]-[7]. However, these methods still suffer from quite restrictive assumptions. In particular, the application of the wideband STFD approach [7] restricts the sliding data window length, whereas the consideration in [6] is limited by

the assumption of linear FM signals with the central frequencies which are known and identical for each source.

In this paper, we obtain a new form of the deterministic ML estimator of the parameters of multiple wideband constant-amplitude polynomial-phase signals received by a sensor array. Our technique is free of any restrictions on the signal waveform parameters and the length of the observation interval. Explicit expressions for the corresponding CRB on the accuracy of estimating the signal DOA and frequency parameters are derived.

Although the presented ML estimator concentrates the problem at hand with respect to the signal nuisance parameters, its computational cost may be still very high, since it involves a nonlinear optimization over the parameter space of a high dimension. Therefore, an approximate solution is considered, originating from the analysis of the ML function in the single chirp case. Using this approximation, we obtain a new form of spatio-temporal matched filter (hereafter referred to as the *chirp beamformer*), which is applicable to the wide class of scenarios with "well-separated" sources that have distinct time-frequency or/and spatial signatures. Our chirp beamforming approach entails solving a 3D optimization problem and, therefore, enjoys essentially simpler implementation than the presented exact ML technique.

Simulation results illustrate the performance of the estimators and validate our CRB analysis.

## 2. SIGNAL MODEL

Assume that  $L$  wideband constant-amplitude polynomial-phase signals impinge on a linear array of  $M$  omnidirectional sensors. The vectors of array outputs obey the following model

$$\mathbf{x}(t) = \mathbf{A}(t)\mathbf{s}(t) + \mathbf{n}(t), \quad t = 0, 1, \dots, N-1 \quad (1)$$

where  $\mathbf{A}(t)$  is the  $M \times L$  time-varying direction matrix,  $\mathbf{s}$  is the  $L \times 1$  vector of wideband nonstationary source waveforms,  $\mathbf{n}(t)$  is the  $M \times 1$  vector of complex circularly Gaussian zero-mean white sensor noise, and  $N$  is the number of snapshots.

The  $l$ th source waveform can be modeled as

$$s_l(t) = \alpha_l e^{j(\omega_{l,0}t + \omega_{l,1}t^2/2 + \dots + \omega_{l,K-1}t^K/K)} = \alpha_l g(\omega_l, t) \quad (2)$$

The work of A.B. Gershman and M. Pesavento was supported by the Natural Sciences and Engineering Research Council (NSERC) of Canada. The work of M.G. Amin was supported by the ONR, Grant N00014-98-1-1076.

where

$$g(\omega_l, t) = \exp \left\{ j \sum_{k=0}^{K-1} \omega_{l,k} \frac{t^{k+1}}{k+1} \right\} \quad (3)$$

$\alpha_l$  is the deterministic complex amplitude,  $\omega_{l,k}$  ( $l = 1, 2, \dots, L$ ;  $k = 0, 1, \dots, K-1$ ) are the unknown frequency parameters, and

$$\tilde{\omega}_l(t) = \sum_{k=0}^{K-1} \omega_{l,k} t^k \quad (4)$$

is the instantaneous frequency of the  $l$ th polynomial-phase waveform. The  $K \times 1$  vector

$$\omega_l = [\omega_{l,0}, \omega_{l,1}, \dots, \omega_{l,K-1}]^T \quad (5)$$

contains the unknown frequency parameters of the  $l$ th signal, and  $K$  is the order of the polynomial-phase model.

The direction matrix

$$\mathbf{A}(t) = [\mathbf{a}(\theta_1, t), \dots, \mathbf{a}(\theta_L, t)]$$

combines the time-varying steering vectors

$$\mathbf{a}(\theta_l, t) = [1, e^{j(\tilde{\omega}_l(t)/c)d_1 \sin \theta_l}, \dots, e^{j(\tilde{\omega}_l(t)/c)d_{M-1} \sin \theta_l}]^T \quad (6)$$

where  $d_i$  is the spacing between the first and the  $(i+1)$ th array sensors. In (6), we assume that the instantaneous signal frequencies  $\tilde{\omega}_l(t)$  ( $t = 1, \dots, L$ ) do not change during the time necessary for a wave to travel across the array aperture<sup>1</sup>. Using (2)-(6), model (1) can be rewritten as

$$\begin{aligned} \mathbf{x}(t) &= \mathbf{A}(\theta, \omega, t) \mathbf{G}(\omega, t) \boldsymbol{\alpha} + \mathbf{n}(t) \\ &= \tilde{\mathbf{A}}(\theta, \omega, t) \boldsymbol{\alpha} + \mathbf{n}(t) \end{aligned} \quad (7)$$

where

$$\boldsymbol{\theta} = [\theta_1, \dots, \theta_L]^T \quad (8)$$

$$\boldsymbol{\omega} = [\omega_1^T, \dots, \omega_L^T]^T \quad (9)$$

$$\boldsymbol{\alpha} = [\alpha_1, \dots, \alpha_L]^T \quad (10)$$

$$\mathbf{G}(\omega, t) = \text{diag}\{g(\omega_1, t), \dots, g(\omega_L, t)\} \quad (11)$$

$$\tilde{\mathbf{A}}(\theta, \omega, t) = \mathbf{A}(\theta, \omega, t) \mathbf{G}(\omega, t) \quad (12)$$

Note that all nuisance parameters (the deterministic source waveforms) are now included in the vector  $\boldsymbol{\alpha}$ .

### 3. DETERMINISTIC ML ESTIMATOR

In this section, we derive the ML estimator of the source DOA's and frequency parameters based on the assumption of deterministic source waveforms. The negative log-likelihood (LL) function is given by

$$\begin{aligned} \mathcal{L}_N(\boldsymbol{\Theta}) &= \sum_{t=0}^{N-1} \|\mathbf{x}(t) - \mathbf{A}(\boldsymbol{\theta}, \boldsymbol{\omega}, t) \mathbf{G}(\boldsymbol{\omega}, t) \boldsymbol{\alpha}\|^2 \\ &= \sum_{t=0}^{N-1} \|\mathbf{x}(t) - \tilde{\mathbf{A}}(\boldsymbol{\theta}, \boldsymbol{\omega}, t) \boldsymbol{\alpha}\|^2 \end{aligned} \quad (13)$$

<sup>1</sup>If necessary, this assumption can be easily relaxed by incorporating into (6) the explicit expression for instantaneous frequency as a function of propagation delay. However, in most of cases, this assumption is valid because the propagation time across the aperture is usually much smaller than the sampling interval.

where the  $(LK + 2L) \times 1$  vector of unknown model parameters is defined as

$$\boldsymbol{\Theta} = [\boldsymbol{\theta}^T, \boldsymbol{\omega}^T, \boldsymbol{\alpha}^T]^T$$

Rewrite (13) as

$$\begin{aligned} \mathcal{L}_N(\boldsymbol{\Theta}) &= \boldsymbol{\alpha}^H \left\{ \sum_{t=0}^{N-1} \tilde{\mathbf{A}}^H(\boldsymbol{\theta}, \boldsymbol{\omega}, t) \tilde{\mathbf{A}}(\boldsymbol{\theta}, \boldsymbol{\omega}, t) \right\} \boldsymbol{\alpha} \\ &\quad - \left\{ \sum_{t=0}^{N-1} \mathbf{x}^H(t) \tilde{\mathbf{A}}(\boldsymbol{\theta}, \boldsymbol{\omega}, t) \right\} \boldsymbol{\alpha} \\ &\quad - \boldsymbol{\alpha}^H \left\{ \sum_{t=0}^{N-1} \tilde{\mathbf{A}}^H(\boldsymbol{\theta}, \boldsymbol{\omega}, t) \mathbf{x}(t) \right\} \\ &\quad + \sum_{t=0}^{N-1} \mathbf{x}^H(t) \mathbf{x}(t) \end{aligned} \quad (14)$$

The minimization of  $\mathcal{L}_N$  over  $\boldsymbol{\alpha}$  yields

$$\hat{\boldsymbol{\alpha}} = \left\{ \sum_{t=0}^{N-1} \tilde{\mathbf{A}}^H(\boldsymbol{\theta}, \boldsymbol{\omega}, t) \tilde{\mathbf{A}}(\boldsymbol{\theta}, \boldsymbol{\omega}, t) \right\}^{-1} \left\{ \sum_{t=0}^{N-1} \tilde{\mathbf{A}}^H(\boldsymbol{\theta}, \boldsymbol{\omega}, t) \mathbf{x}(t) \right\}$$

Substituting this expression into (14), we obtain the negative concentrated LL function

$$\begin{aligned} \mathcal{L}_N(\boldsymbol{\theta}, \boldsymbol{\omega}) &= \sum_{t=0}^{N-1} \mathbf{x}^H(t) \mathbf{x}(t) - \left\{ \sum_{t=0}^{N-1} \mathbf{x}^H(t) \tilde{\mathbf{A}}(\boldsymbol{\theta}, \boldsymbol{\omega}, t) \right\} \\ &\quad \times \left\{ \sum_{t=0}^{N-1} \tilde{\mathbf{A}}^H(\boldsymbol{\theta}, \boldsymbol{\omega}, t) \tilde{\mathbf{A}}(\boldsymbol{\theta}, \boldsymbol{\omega}, t) \right\}^{-1} \\ &\quad \times \left\{ \sum_{t=0}^{N-1} \tilde{\mathbf{A}}^H(\boldsymbol{\theta}, \boldsymbol{\omega}, t) \mathbf{x}(t) \right\} \end{aligned} \quad (15)$$

Ignoring the constant terms, the positive concentrated LL function is then given by

$$\begin{aligned} \mathcal{L}_P(\boldsymbol{\theta}, \boldsymbol{\omega}) &= \left\{ \sum_{t=0}^{N-1} \mathbf{x}^H(t) \tilde{\mathbf{A}}(\boldsymbol{\theta}, \boldsymbol{\omega}, t) \right\} \\ &\quad \times \left\{ \sum_{t=0}^{N-1} \tilde{\mathbf{A}}^H(\boldsymbol{\theta}, \boldsymbol{\omega}, t) \tilde{\mathbf{A}}(\boldsymbol{\theta}, \boldsymbol{\omega}, t) \right\}^{-1} \\ &\quad \times \left\{ \sum_{t=0}^{N-1} \tilde{\mathbf{A}}^H(\boldsymbol{\theta}, \boldsymbol{\omega}, t) \mathbf{x}(t) \right\} \end{aligned} \quad (16)$$

The ML estimator

$$[\hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\omega}}] = \arg \max_{\boldsymbol{\theta}, \boldsymbol{\omega}} \mathcal{L}_P(\boldsymbol{\theta}, \boldsymbol{\omega}) \quad (17)$$

jointly estimates the direction and the frequency parameters  $\boldsymbol{\theta}$  and  $\boldsymbol{\omega}$ , respectively. It requires a highly nonlinear optimization of the function (16) over these variables.



#### 4. CRAMÉR-RAO BOUND

In this section, we present closed-form expressions for the exact CRB on the accuracy of estimating the signal model parameters.

*Theorem 1:* Let the observations (7) satisfy the following statistical model:

$$\mathbf{x}(t) \sim \mathcal{N}(\tilde{\mathbf{A}}(\theta, \omega, t)\alpha, \sigma^2 \mathbf{I}) \quad (18)$$

Then, the Fisher Information Matrix (FIM) is given by

$$\mathbf{F} = \begin{bmatrix} \mathcal{F} & \mathbf{0} \\ \mathbf{0}^T & F_{\sigma^2 \sigma^2} \end{bmatrix} \quad (19)$$

where

$$\mathcal{F} = \begin{bmatrix} \mathbf{F}_{\theta\theta} & \mathbf{F}_{\theta\alpha} & \mathbf{F}_{\theta\nu_0} & \cdots & \mathbf{F}_{\theta\nu_{K-1}} \\ \mathbf{F}_{\theta\alpha}^T & \mathbf{F}_{\alpha\alpha} & \mathbf{F}_{\alpha\nu_0} & \cdots & \mathbf{F}_{\alpha\nu_{K-1}} \\ \mathbf{F}_{\theta\nu_0}^T & \mathbf{F}_{\alpha\nu_0}^T & \mathbf{F}_{\nu_0\nu_0} & \cdots & \mathbf{F}_{\nu_0\nu_{K-1}} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{F}_{\theta\nu_{K-1}}^T & \mathbf{F}_{\alpha\nu_{K-1}}^T & \mathbf{F}_{\nu_0\nu_{K-1}} & \cdots & \mathbf{F}_{\nu_{K-1}\nu_{K-1}} \end{bmatrix}$$

$$\mathbf{F}_{\theta\theta} = \frac{2}{\sigma^2} \text{Re} \left\{ \sum_{t=0}^{N-1} \Delta^H \tilde{\mathbf{D}}^H(\theta, \omega, t) \tilde{\mathbf{D}}(\theta, \omega, t) \Delta \right\}$$

$$\mathbf{F}_{\theta\alpha} = \frac{2}{\sigma^2} \text{Re} \left\{ \sum_{t=0}^{N-1} \Delta^H \tilde{\mathbf{D}}^H(\theta, \omega, t) \tilde{\mathbf{A}}(\theta, \omega, t) \mathbf{Q} \right\}$$

$$\mathbf{F}_{\theta\nu_k} = \frac{2}{\sigma^2} \text{Re} \left\{ \sum_{t=0}^{N-1} \Delta^H \tilde{\mathbf{D}}^H(\theta, \omega, t) \times \{T_k(t) \odot (\tilde{\mathbf{A}}(\theta, \omega, t) \Delta)\} \right\}$$

$$\mathbf{F}_{\alpha\alpha} = \frac{2}{\sigma^2} \text{Re} \left\{ \sum_{t=0}^{N-1} \mathbf{Q}^H \tilde{\mathbf{A}}^H(\theta, \omega, t) \tilde{\mathbf{A}}(\theta, \omega, t) \mathbf{Q} \right\}$$

$$\mathbf{F}_{\alpha\nu_k} = \frac{2}{\sigma^2} \text{Re} \left\{ \sum_{t=0}^{N-1} \mathbf{Q}^H \tilde{\mathbf{A}}^H(\theta, \omega, t) \times \{T_k(t) \odot (\tilde{\mathbf{A}}(\theta, \omega, t) \Delta)\} \right\}$$

$$\mathbf{F}_{\nu_k\nu_m} = \frac{2}{\sigma^2} \text{Re} \left\{ \sum_{t=0}^{N-1} \left\{ (\Delta^H \tilde{\mathbf{A}}^H(\theta, \omega, t)) \odot T_k^H(t) \right\} \times \{T_m(t) \odot (\tilde{\mathbf{A}}(\theta, \omega, t) \Delta)\} \right\}$$

$$F_{\sigma^2 \sigma^2} = \frac{NM}{\sigma^4}$$

$$\mathbf{F}_{\theta\sigma^2} = \mathbf{F}_{\alpha\sigma^2} = \mathbf{F}_{\nu_k\sigma^2} = \mathbf{0}$$

$$\Delta = \text{diag}\{\alpha_1, \dots, \alpha_L\}$$

$$\tilde{\alpha} = [\text{Re}\{\alpha\}^T, \text{Im}\{\alpha\}^T]^T$$

$$\tilde{\mathbf{D}}(\theta, \omega, t) = \mathbf{D}(\theta, \omega, t) \mathbf{G}(\omega, t)$$

$$\mathbf{D}(\theta, \omega, t) = \left[ \frac{\partial \mathbf{a}(\theta_1, \omega_1, t)}{\partial \theta_1}, \dots, \frac{\partial \mathbf{a}(\theta_L, \omega_L, t)}{\partial \theta_L} \right]$$

$$\mathbf{Q} = [\mathbf{I}, j\mathbf{I}]$$

$$T_k(t) = jt^k \left( \frac{t}{k+1} \mathbf{E} + \mathbf{u} \mathbf{c}^T \right)$$

$$\mathbf{u} = \frac{1}{c} [0, d_1, \dots, d_{M-1}]^T$$

$$\mathbf{c} = [\cos \theta_1, \cos \theta_2, \dots, \cos \theta_L]^T$$

$\mathbf{E}$  is the matrix containing ones in all positions,  $\mathbf{0}$  is the vector of zeros,

$$\begin{aligned} \boldsymbol{\nu} &= \text{vec}\{\boldsymbol{\Omega}^T\} \\ &= [\nu_0^T, \dots, \nu_{K-1}^T]^T \end{aligned} \quad (20)$$

$$\boldsymbol{\Omega} = \begin{bmatrix} \omega_{1,0} & \omega_{2,0} & \cdots & \omega_{L,0} \\ \omega_{1,1} & \omega_{2,1} & \cdots & \omega_{L,1} \\ \vdots & \vdots & \cdots & \vdots \\ \omega_{1,K-1} & \omega_{2,K-1} & \cdots & \omega_{L,K-1} \end{bmatrix} \quad (21)$$

and  $\text{vec}\{\cdot\}$  represents the so-called vectorization operator stacking the columns of a matrix to form a column vector.

*Proof:* See [8].

It is important to stress that vector (20) contains the same signal frequency parameters as those included in  $\boldsymbol{\omega}$ . However, these parameters are ordered in a different way. To clarify the difference between  $\boldsymbol{\omega}$  and  $\boldsymbol{\nu}$ , note that

$$\boldsymbol{\omega} = \text{vec}\{\boldsymbol{\Omega}\} \quad (22)$$

In Theorem 1, we use the vector  $\boldsymbol{\nu}$  rather than  $\boldsymbol{\omega}$  for the sake of mathematical convenience, since the CRB derivation in terms of  $\boldsymbol{\nu}$  leads to simpler expressions for the FIM subblocks.

#### 5. CHIRP BEAMFORMER

The associated computational cost of the ML estimator (16)-(17) may not be always acceptable. In this section, we simplify the ML estimator by deriving the so-called *chirp beamformer* which requires a simpler 3D search instead of global optimization. Assuming the single source case<sup>2</sup>, we rewrite the LL function (16) as

$$\mathcal{L}_P(\theta_1, \omega_1) = \frac{1}{NM} \left| \sum_{t=0}^{N-1} \mathbf{x}^H(t) \tilde{\mathbf{a}}(\theta_1, \omega_1, t) \right|^2 \quad (23)$$

where

$$\tilde{\mathbf{a}}(\theta_1, \omega_1, t) = g(\omega_1, t) \mathbf{a}(\theta_1, \omega_1, t) \quad (24)$$

and the property  $\tilde{\mathbf{a}}^H \tilde{\mathbf{a}} = M$  is used. Assuming a chirp signal, we have  $\omega_1 = [\omega_{1,0}, \omega_{1,1}]^T$  and, hence, there are only three parameters  $\{\theta_1, \omega_{1,0}, \omega_{1,1}\}$ , which correspond to the DOA, frequency, and the chirp rate, respectively.

Introducing the simplified (subscript-free) notation

$$\theta = \theta_1, \quad \xi = \omega_{1,0}, \quad \zeta = \omega_{1,1} \quad (25)$$

and omitting the constant factor  $1/M$ , we can rewrite the right-hand side of (23) as the following function:

$$f(\theta, \xi, \zeta) = \frac{1}{N} \left| \sum_{t=0}^{N-1} \mathbf{x}^H(t) \tilde{\mathbf{a}}(\theta, \xi, \zeta, t) \right|^2 \quad (26)$$

<sup>2</sup>This assumption will be relaxed later.

The function (26) is referred to as the chirp beamformer<sup>3</sup>.

The parameters of interest can be obtained from the main maxima of (26) by means of a 3D search over the variables  $\{\theta, \xi, \zeta\}$ . The chirp beamformer (26) can be easily applied to the multiple source case under the condition that the sources are "well-separated" in one or more parameters in (25). This property follows from the structure of (26), which is linear with respect to the second-order moments of  $\mathbf{x}$ . Therefore, as in the case of the conventional beamformer [9], [10] which is widely used in narrowband array processing, the chirp beamformer (26) can be straightforwardly extended to the multiple source case.

Interestingly, the chirp beamformer has quite a different structure as compared to the conventional beamformer. The latter function is given by [10]

$$\begin{aligned} f_{CB}(\theta) &= \mathbf{a}^H(\theta) \hat{\mathbf{R}} \mathbf{a}(\theta) \\ &= \frac{1}{N} \sum_{t=0}^{N-1} |\mathbf{x}^H(t) \mathbf{a}(\theta)|^2 \end{aligned} \quad (27)$$

where

$$\hat{\mathbf{R}} = \frac{1}{N} \sum_{t=0}^{N-1} \mathbf{x}(t) \mathbf{x}^H(t) \quad (28)$$

is the sample covariance matrix, and the steering vector does not depend on the temporal index  $t$ . Comparing (26) and (27), we maintain that the conventional beamformer represents the *sum of the squared absolute values* of vector inner products, whereas the chirp beamformer, on the other hand, is determined by the *squared absolute value of the sum of inner products*. This essential difference between (26) and (27) can be explained by the fact that in the chirp signal case, the signal temporal characteristics are taken into account by means of the parametric time-domain polynomial-phase model. Obviously, this corresponds to the so-called *coherent time-domain processing*, whereas in the conventional narrowband case the snapshots  $\mathbf{x}(t)$  are assumed to be independent and, therefore, the processing in (27) remains *incoherent* in time-domain.

An interesting relationship between the chirp beamformer (26) and the traditional estimation techniques can be obtained for the conventional harmonic signal case ( $\zeta = 0$ ). In this case, we have

$$\tilde{\mathbf{a}}(\theta, \xi, \zeta, t) = \tilde{\mathbf{a}}(\theta, \xi, t) = e^{j\xi t} \mathbf{a}(\theta, \xi) \quad (29)$$

where the vector  $\mathbf{a}(\theta, \xi)$  is the conventional steering vector, which coincides to that in (27). Hence, the beamforming function (26) can be transformed to

$$\begin{aligned} f(\theta, \xi) &= |\mathbf{X}^H(\xi) \mathbf{a}(\theta, \xi)|^2 \\ &= \mathbf{a}^H(\theta, \xi) \mathbf{X}(\xi) \mathbf{X}^H(\xi) \mathbf{a}(\theta, \xi) \end{aligned} \quad (30)$$

where

$$\mathbf{X}(\xi) = \frac{1}{\sqrt{N}} \sum_{t=0}^{N-1} \mathbf{x}(t) e^{-j\xi t} \quad (31)$$

<sup>3</sup>We use this term because of the obvious analogy with the narrowband conventional beamformer [10] which can be easily derived from the conventional deterministic ML estimator under the single-source assumption [11].

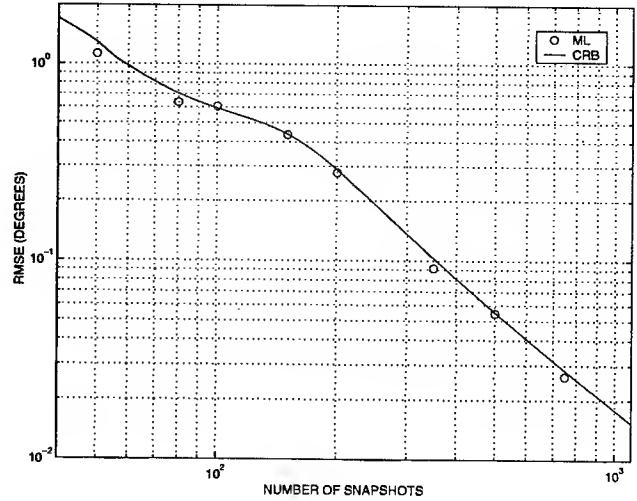


Figure 1: Comparison of the DOA estimation RMSE of the ML estimator and the CRB versus the number of snapshots.

is the  $M \times 1$  vector of the Fourier-transformed array outputs. The estimator (30) represents a single-snapshot variant of the frequency-domain conventional beamformer [9]

$$f_{CB}(\theta, \xi) = \mathbf{a}^H(\theta, \xi) \hat{\mathbf{R}}(\xi) \mathbf{a}(\theta, \xi) \quad (32)$$

where

$$\hat{\mathbf{R}}(\xi) = \frac{1}{P} \sum_{\tau=0}^{P-1} \mathbf{X}(\xi, \tau) \mathbf{X}^H(\xi, \tau) \quad (33)$$

is the sample spectral density matrix, the time index  $\tau$  determines the location of the respective short Fourier transform sliding window<sup>4</sup>, and  $P$  is the total number of sliding windows (or, in the other words, the number of frequency-domain snapshots).

Similarly to the chirp beamformer (26), a *polynomial-phase beamformer* can be defined that corresponds to a more general polynomial-phase signal model. In this case, the number of parameters in (26) will increase, depending on the polynomial-phase model order.

## 6. SIMULATIONS

In all examples, we assume a uniform linear array (ULA) with the half-wavelength spacing. In the first example, we assume a ULA of  $M = 10$  sensors which receives two equipowered chirp sources with SNR = 0 dB and DOA's  $\theta_1 = 10^\circ$  and  $\theta_2 = 15^\circ$  relative to the broadside. The sources have the following frequency parameters:  $\omega_{1,0} = 1.2566$ ,  $\omega_{1,1} = -0.0151$ ,  $\omega_{2,0} = 0.0628$ , and  $\omega_{2,1} = 0.0151$ . In Fig. 1, the DOA estimation RMSE of the ML estimator (16)-(17) and the theoretically obtained direction estimation CRB versus the number of snapshots  $N$  are shown.

<sup>4</sup>This index is not shown in (30) because it is a particular case where the single window, whose length is equal to the whole observation length, is used.

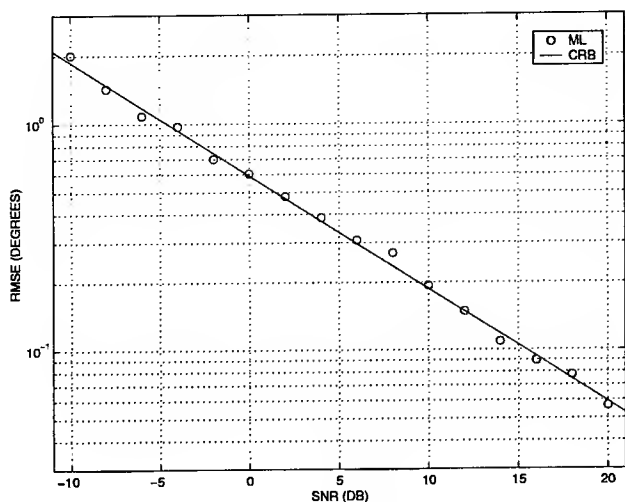


Figure 2: Comparison of the DOA estimation RMSE of the ML estimator and the CRB versus the SNR.

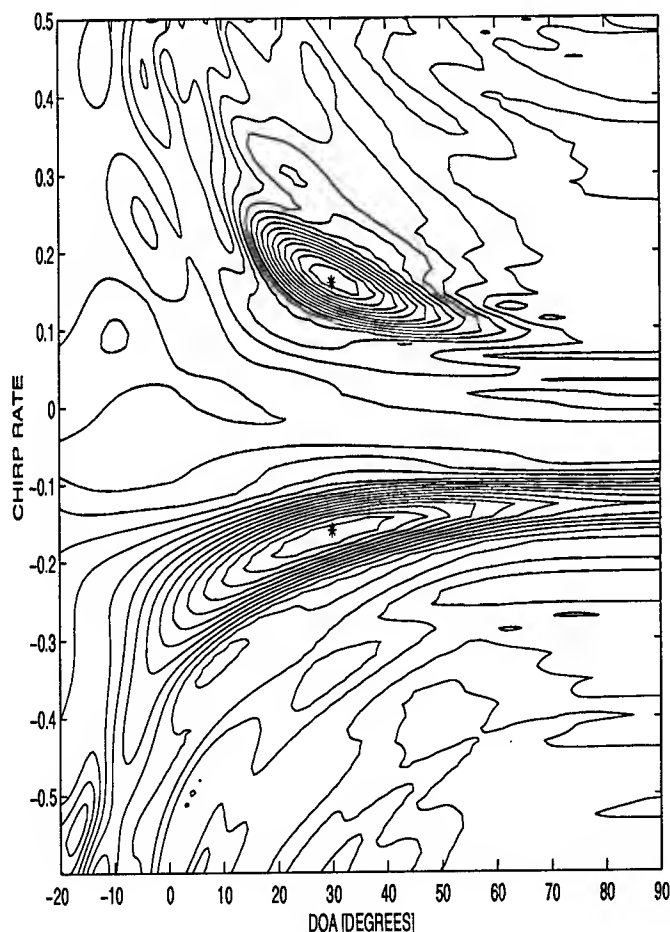


Figure 3: 2D slice of the chirp beamformer. The true source locations are indicated by stars.

In our second example, the same parameters are used except for SNR and  $N$ . We assume that  $N = 100$  and the performance is examined versus the SNR. The RMSE of the ML estimator and the CRB are displayed in Fig. 2.

In our third example, we assume a ULA of  $M = 15$  sensors and two equi-powered chirp sources with the SNR = 0 dB and  $\theta_1 = \theta_2 = 30^\circ$ . The following parameters are used:  $N = 15$ ,  $\omega_{1,0} = \omega_{2,0} = 0.8000$ ,  $\omega_{1,1} = -0.1600$ , and  $\omega_{2,1} = 0.1600$ . Fig. 3 displays the 2D slice of the 3D chirp beamforming function evaluated at  $\xi = 0.8000$ . From this figure, we observe that the chirp beamformer is able to resolve closely spaced sources (and even sources having the same DOA's and initial frequencies), based solely on the difference of their chirp rates.

## REFERENCES

- [1] S. Peleg and B. Porat, "Estimation and classification of polynomial-phase signals," *IEEE Trans. Inform. Theory*, vol. 37, pp. 422-430, March 1991.
- [2] S. Shamsunder, G. Giannakis, and B. Friedlander, "Estimating random amplitude polynomial phase signals: A cyclostationary approach," *IEEE Trans. Signal Processing*, vol. 43, pp. 492-505, Feb. 1995.
- [3] O. Besson, M. Ghogho, and A. Swami, "Parameter estimation for random amplitude chirp signals," *IEEE Trans. Signal Processing*, vol. 47, pp. 3208-3219, Dec. 1999.
- [4] A. Zeira and B. Friedlander, "Joint direction finding, signal, and channel response estimation for a polynomial phase signal in a multipath channel," in *Proc. 30th Asilomar Conf. Sign., Syst., Comp.*, pp. 733-737, Nov. 1997.
- [5] M.G. Amin, "Spatial time-frequency distributions for direction finding and blind source separation," *Proc. SPIE*, vol. 3723, pp. 62-70, 1999.
- [6] B. Völcker and B. Ottersten, "Linear chirp parameter estimation from multi channel data," in *Proc. 32nd Asilomar Conf. Sign., Syst., Comp.*, Nov. 1999.
- [7] A.B. Gershman and M.G. Amin, "Wideband direction of arrival estimation of multiple chirp signals using spatial time-frequency distributions," *IEEE Signal Processing Letters*, vol. 7, June 2000.
- [8] A.B. Gershman, M. Pesavento, and M.G. Amin, "Estimating parameters of multiple wideband polynomial-phase sources in sensor arrays," submitted.
- [9] J.F. Böhme, "Array processing," in *Advances in Spectrum Estimation*, S. Haykin, Ed., Prentice-Hall, Englewood Cliffs, NJ, 1991, vol. II, pp. 1-63.
- [10] H. Krim and M. Viberg, "Two decades of array signal processing research: The parametric approach," *IEEE Signal Processing Magazine*, vol. 13, pp. 67-94, July 1996.
- [11] V. Katkovnik and A.B. Gershman, "A local polynomial approximation based beamforming for source localization and tracking in nonstationary environments," *IEEE Signal Processing Letters*, vol. 7, pp. 3-5, Jan. 2000.

# On the Use of Space-Time Adaptive Processing and Time-Frequency Data Representations for Detection of Near-Stationary Targets in Monostatic Clutter

*D. C. Braunreiter, H.-W. Chen, M. L. Cassabaum, J. G. Riddle, A. A. Samuel, J. F. Scholl and H. A. Schmitt*

Raytheon Missile Systems  
P. O. Box 11337  
Tucson, Arizona 85734, USA

## ABSTRACT

The detection of near-stationary targets in mainlobe clutter is a problem that has recently generated a great deal of interest within the Department of Defense community. Some examples of these types of targets are surface vehicles, missile launchers and loitering (micro-) Unmanned Aerial Vehicles (UAVs). The root of the difficulty lies in the fact that conventional radar processing loses the ability to use the Doppler of the target to discriminate it from the clutter. Indeed, the target need not even be nearly stationary for this to be a problem – even a rapidly moving target can exhibit low Doppler if its velocity vector is nearly perpendicular to the velocity vector of the observation platform. Raytheon Systems Company (Raytheon) has been investigating a number of advanced algorithmic solutions to this problem within the context of providing a dual-mission capability to currently fielded RF missile systems. This paper describes a processing architecture that combines preprocessing, Time-Frequency Transforms and Best Bases algorithms and discusses some preliminary results.

## 1. INTRODUCTION

We propose a novel method for the detection of stationary targets in monostatic clutter [1, 2]. This is a region where conventional Space-Time Adaptive Processing (STAP) algorithms experience difficulty since there is no longer any Doppler discriminant

available. While a number of hardware solutions have been proposed, for example, the addition of an adjunct infrared sensor and associated processing hardware, an RF-based algorithmic solution remains a very attractive option. This option should also provide the basis for a dual-mission RF missile, thereby extending the capability of currently fielded hardware. Raytheon is investigating a number of algorithmic approaches to this problem; in this manuscript, we concentrate on the use of Time-Frequency Transforms in combination with various pre-filtering and post-processing algorithms. We have achieved the best performance by first preprocessing the data using whitening or Wiener filters, then mapping the 1-D time series data onto a 2-D Time-Frequency image using a Wigner-Ville Transform (WVT) to enhance features, and finally employing a Best Bases type of algorithm for feature extraction. We note that our approach to this problem is similar in spirit to that proposed by Haykin in References 3-4.

The proposed target detection algorithm is shown schematically in Figure 1; the red outlined area indicates the nonstandard processing portion of this algorithm. Notice that the Time-Frequency Analysis (TFA)-Best Bases processing stream allows the natural introduction of feature fusion. This is an important characteristic, since a number of programs at Raytheon have had considerable success using feature fusion for improving target classification.

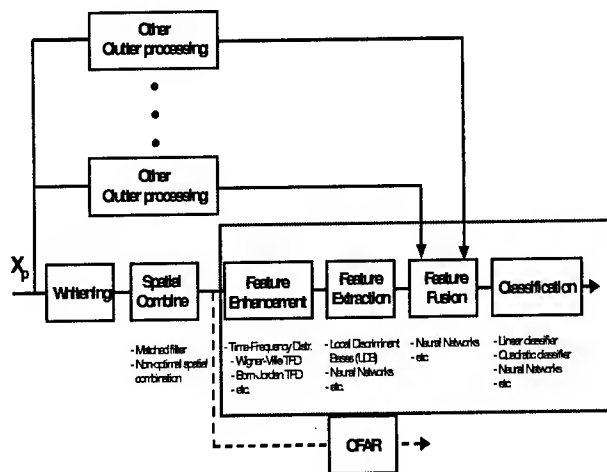


Figure 1: Time-Frequency Detection Scheme

Figure 2 below shows the output of the Feature Extraction block in Figure 1 for 0 dBsm target. At this point the data has been filtered and passed through a WVT. The Best Bases algorithm that we have used in this analysis is the Local Discriminant Bases (LDB) [5] algorithm of Coifman and Saito. LDB was originally developed in 1994 as a technique for analyzing object classification problems. Since then, extensions have been developed for regression, optimization and signal de-mixing applications.

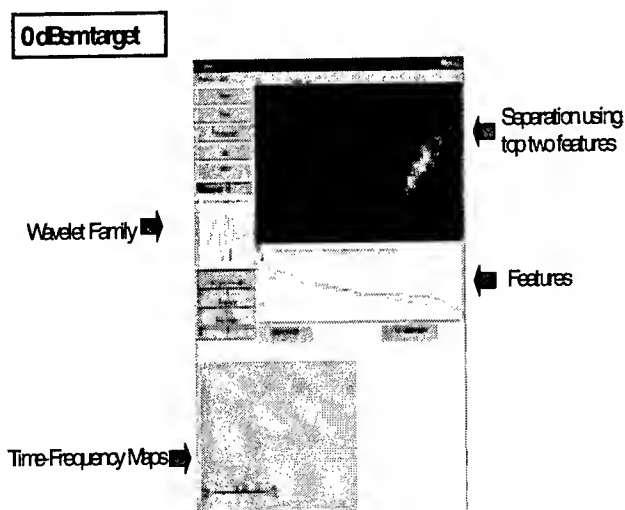


Figure 2: LDB Screen for Target Detection

## 2. TIME-FREQUENCY ANALYSIS USING RANGE-DOPPLER MAP PHASES

Raytheon has also been investigating the use of phase information from the Range-Doppler maps for radar signals. In conventional monopulse seekers, only the amplitudes of the complex-valued range/Doppler-filter outputs are used for target detection and/or identification - the random-like phases are seldom considered helpful. However, it has recently been proposed [6] that it is essential to utilize the whole complex-valued range-Doppler image because much of the information about the target is contained in the phase. We have used simple correlation and spectrum estimation techniques to extract the phase signals, and used some simple detrending techniques to correct the clutter leakage of the Doppler FFT.

Further improved results are expected when more advanced spectrum estimation and FFT leakage correction techniques are employed. For example, the current Power Spectral Density (PSD) technique can only detect targets at different range-gates. We do not know the target Doppler, and cannot resolve targets located at the same range-gate but with different Doppler frequencies. Furthermore, we can not distinguish different targets that may have similar PSD functions. We have begun to investigate Time-frequency analysis for this problem and believe that it has promise here since it provides an ability to measure the whole frequency components at different Doppler frequencies.

As shown in the PSD plots in Figures 3, 4, and 5, the phase signals with targets have much higher low frequency components than the phase signal with clutter only. Therefore, we can detect targets at the range-gate of the signal using the low frequency components. However, we do not know the target Doppler and cannot discriminate among different targets.

As shown in the two-dimensional Wigner-Ville plot (*c.f.*, Figure 3), the phase signal with clutter only has a wide frequency band across almost the whole Doppler duration. There are also two weak linear chirps appearing in the lower Doppler. However, when a target T-60 is included in the RF

signal, the energy at the lower Doppler (where the target Doppler located) is much lower than the signal with clutter only, as shown in Figure 4. For a different target T-120, we can find the similar result as shown in Figure 5. It is interesting to note that this target generates two linear chirps at the higher Doppler; therefore, time-frequency analysis may help discriminate among different targets.

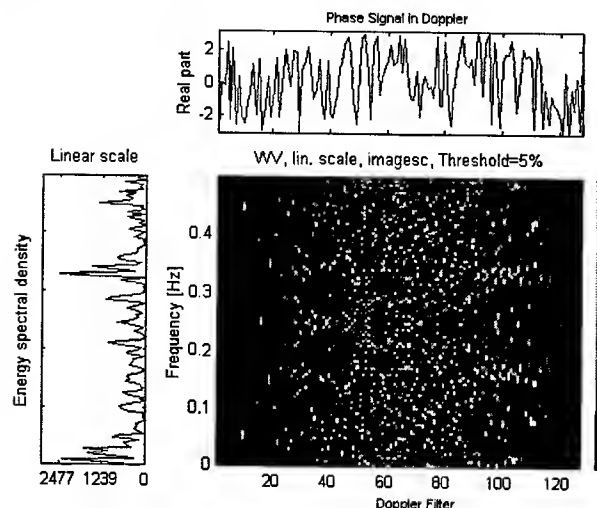


Figure 3. Clutter Only Phase Signal

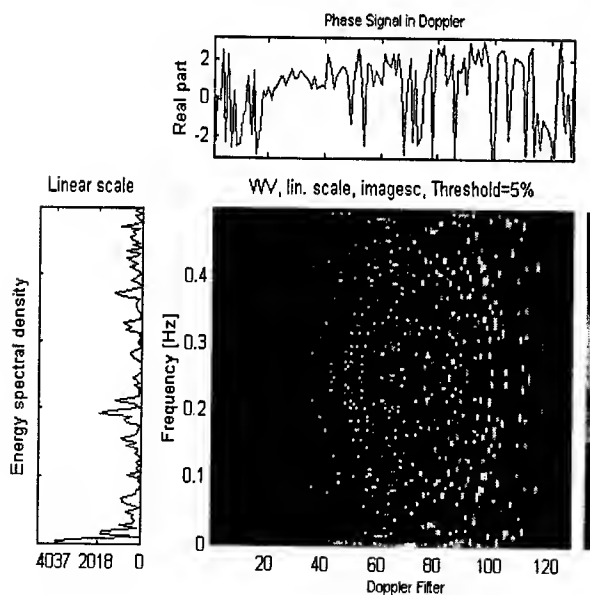


Figure 4. Phase Signal with Clutter, Receiver Noise and Target T-60

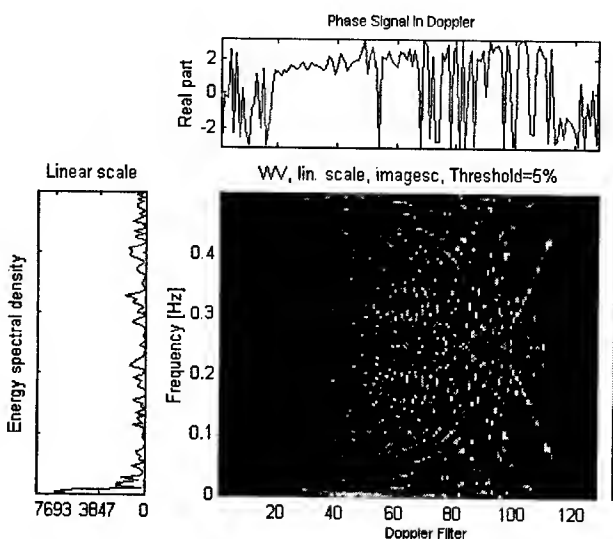


Figure 5. Phase Signal with Clutter, Receiver Noise and Target T-120

### 3. CONTINUOUS WAVELET TRANSFORMS

Raytheon has also been investigating the use of continuous wavelet transforms (CWT) for target detection and feature extraction. The advantage of using this method is that there are no artificial interference terms generated in the analysis unlike that of the WVT, and there are a larger number of possible "mother" wavelet functions from which we can obtain more optimal time-frequency analyses with. The downside of this scheme is the high computational complexity. We are currently investigating means to improve the CWT either by implementing further improvements to the algorithm itself, and/or by performing the calculations using fast analog signal processors [7]. Of course, the WVT can be implemented on these analog devices as well.

### 4. CONCLUSION

In this manuscript, we have presented some preliminary result of using TFT, in combination with pre-filtering and post-processing, to detect near stationary targets in main lobe clutter. Raytheon is also investigating a number of other algorithms, including polarization STAP, covariance matrix conditioning, waveform diversity, non-decimated wavelet transforms and higher order statistics. We have also begun looking at some very interesting work on optimized kernel TFTs [8, 9].

## 5. REFERENCES

- [1] A. A. Samuel, H. A. Schmitt, G. T. David, H.-W. Chen, and D. C. Braunreiter, "On the Use of Space-time Adaptive Processing and Multi-resolution Data Representations for the Detection of Near-Stationary Targets in Monostatic Clutter," 45<sup>th</sup> Tri-Services Radar Conference, Naval Postgraduate School, Monterey, CA, June, 1999.
- [2] H.-W. Chen, H. A. Schmitt, G. T. David, D. C. Braunreiter, A. A. Samuel and Dennis. M. Healy, "Detection and Direction Estimation of Near-Stationary Targets in Monostatic Clutter Using RF Phases," 46<sup>th</sup> Tri-Services Radar Conference, U.S. Air Force Academy, Colorado Springs, CO, June, 2000.
- [3] S. Haykin and T. Bhattacharya, "Wigner-Ville Distribution: An Important Functional Block for Radar Target Detection in Clutter," Twenty-Eighth Asilomar Conference on Signals, Systems and Computers, 1, Volume 1, 1994.
- [4] S. Haykin and D. J. Thomson, "Signal Detection in a Nonstationary Environment Reformulated as an Adaptive Pattern Classification Problem," Proceedings of the IEEE, 86, no. 11, pp. 2325-2344, 1998.
- [5] N. Saito and R. Coifman, 'Local Discriminant Bases,' in Mathematical Imaging: Wavelet Applications in Signal and Image Processing, A. F. Laine and M. A. Unser, eds., Proc. SPIE 2303, pp. 2-14, 1994.
- [6] A. Rihaczek and S. Hershkowitz, *Radar Resolution and Complex-Image Analysis*, Artech House, 1996.
- [7] J. McElvain, J. D. Langan, R. Behm, M. Costolo, and A. J. Heeger, "Spatial Frequency Filtering Using Hybrid Polymer/VLSI Technology," preprint, (1999).
- [8] R. G. Baraniuk and D. L. Jones, "Signal-Dependent Time-Frequency Analysis Using a Radially Gaussian Kernel," Signal Processing, 32, no. 3, pp. 263-284, 1993.
- [9] D. L. Jones and R. G. Baraniuk, "An Adaptive Optimal-Kernel Time-Frequency Representation," IEEE Transactions on Signal Processing, 43, no. 11, pp. 2361-2371, 1995.

# APPLICATION OF ADAPTIVE JOINT TIME-FREQUENCY PROCESSING TO ISAR IMAGE FORMATION

Hao Ling and Junfei Li

Dept. of Electrical and Computer Engineering  
The Univ. of Texas at Austin  
Austin, TX 78712-1084, USA

## ABSTRACT

Two applications of the adaptive joint time-frequency (AJTF) algorithm for ISAR image formation are presented. First, AJTF is utilized for ISAR motion estimation and compensation. Focused images from measured radar data are presented to illustrate the effectiveness of the algorithm when applied to in-flight aircraft data. Second, the AJTF algorithm is extended to detect the presence of chaotic, three-dimensional motions in an articulating target. Preliminary test results on measured data show that the algorithm can correctly detect those imaging intervals where significant three-dimensional motions exist.

## 1. INTRODUCTION

High-resolution inverse synthetic aperture radar (ISAR) imaging is a promising tool for non-cooperative target identification (NCTI). The main challenge in ISAR-based NCTI is to form a well-focused image of an articulating target with unknown motion. In this paper, we first review the application of joint time-frequency methods for ISAR image formation. By using an adaptive joint time-frequency (AJTF) algorithm to estimate the phase of the prominent scatterers, we show that the target motion can be estimated and a focused image of the target can be constructed. Results of applying the algorithm to measured ISAR data are presented and discussed. Secondly, we report on our recent work to extend the AJTF algorithm to address the more challenging situation when the motion of the target is not limited to a two-dimensional plane. In particular, we discuss our research to detect the presence of three-dimensional motion using the AJTF algorithm.

## 2. ISAR MOTION COMPENSATION USING JOINT TIME-FREQUENCY ALGORITHM

We first review the application of joint time-frequency methods for ISAR image formation. To form a focused image from raw radar data, it is customary to first carry out a coarse alignment of the data in the range dimension,

followed by fine motion compensation in the cross range dimension. Joint time-frequency techniques have been shown to be a useful tool to carry out the fine motion compensation [1,2]. We assume that after the coarse range alignment, all the scatterers are located in their respective range cells. The radar backscattered signal as a function of dwell time  $t$  in a particular range cell can be written as

$$E(t) = \sum_{k=1}^N A_k \exp[-j \frac{4\pi f}{c} (R(t) + x_k \cos\theta(t) + y_k \sin\theta(t))] \quad (1)$$

where  $N$  is the number of point scatterers in that range cell, and  $A_k$ ,  $x_k$ ,  $y_k$  are respectively the scattering amplitude, down range position and cross range position of the  $k^{\text{th}}$  point scatterer.  $R(t)$  is the residual uncompensated translation displacement and  $\theta(t)$  is the rotational displacement. Due to translation and rotational motion, the Doppler frequency versus dwell time behavior of the point scatterers within this range cell is not constant in the joint time-frequency plane (see Fig. 1). An effective JTF technique to extract the motion parameters is based on a search and projection procedure to represent the phase behavior of the signal  $E(t)$ . This procedure is based on the adaptive spectrogram proposed in [3], and is similar in concept to a one-term matching pursuit algorithm [4]. We shall term it the adaptive JTF (AJTF) algorithm. To find the motion parameters, basis functions in the form of

$$h(t) = \exp[-j(a_1 t + a_2 t^2 + a_3 t^3)] \quad (2)$$

are chosen. We search for the basis function over the parameter space  $(a_1, a_2, a_3)$  that best represents the time-frequency behavior of the signal by maximizing the projection of the signal onto the basis:

$$\max_{a_1, a_2, a_3} \left| \int E(t) h^*(t) dt \right|^2 \quad (3)$$

After the time-varying phase for the strongest point scatterer is found, we multiply the original signal by the conjugate of this phase factor to compensate for the translation motion. This algorithm can also be extended to



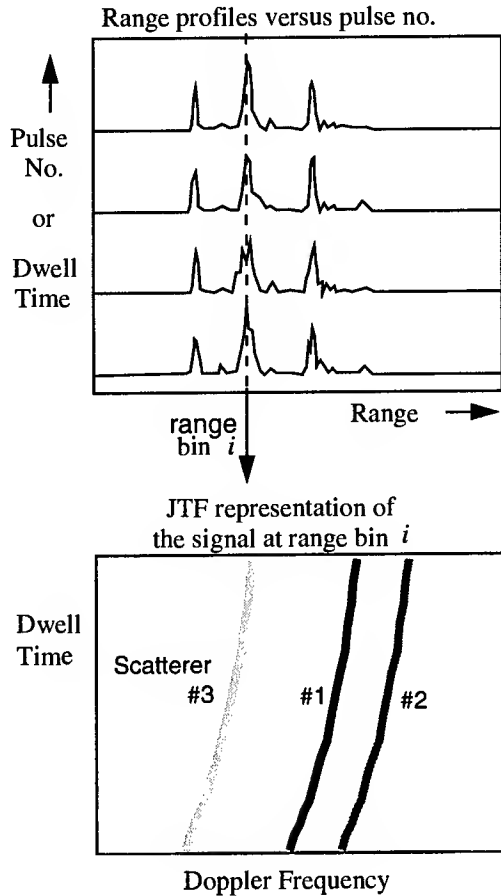


Fig. 1. Fine motion compensation is carried out by the Doppler frequency versus dwell time behavior of the strong point scatterer in the signal.

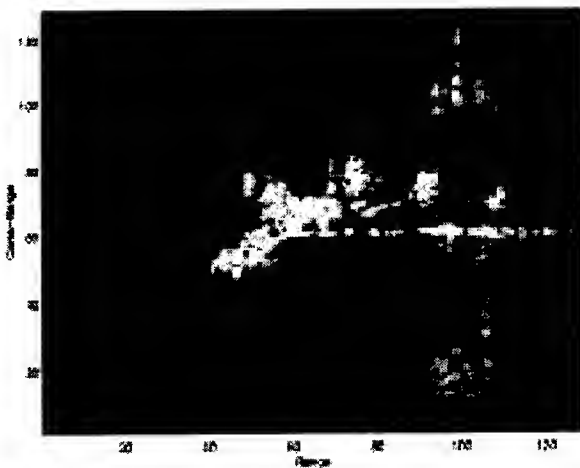


Fig. 2. ISAR image of an in-flight aircraft obtained after AJTF motion compensation.

multiple range cells to correct for higher-order rotation motion. After applying the JTF motion compensation, the standard FFT processing in the dwell time domain brings the signal into the cross range image domain. Fig. 2 shows an example of applying the AJTF algorithm to measured ISAR data of an in-flight aircraft. The shape of the aircraft is clearly visible in the resulting image after the AJTF motion compensation.

### 3. THREE-DIMENSIONAL MOTION DETECTION USING JOINT TIME-FREQUENCY ALGORITHM

One basic assumption of standard motion compensation algorithms is that the target only undergoes motion in a two-dimensional plane during the dwell duration needed to form an image. From several independent examinations of measured ISAR data sets recently, it was reported that the presence of three-dimensional motion is quite detrimental to focusing the image [5-7]. We shall report on our recent work to extend the AJTF algorithm to address the more challenging situation when the motion of the target is not limited to a two-dimensional plane. In particular, we discuss our research to detect the presence of three-dimensional motion using the AJTF algorithm.

Allowing for arbitrary three-dimensional motion in space, we consider the following model as a generalization of the model for two-dimensional motion in (1):

$$E(t) = \sum_{k=1}^N A_k \exp[-j \frac{4\pi f_c}{c} (x_k + y_k \theta + z_k \phi)] \quad (4)$$

where  $\theta$  is the azimuth angle of the target with respect to the radar, and  $\phi$  is the elevation angle. In (4), it is assumed that the translation motion has been removed and that the standard small-angle, small bandwidth approximations apply. This model reduces to the standard two-dimensional motion model when  $\theta$  and  $\phi$  are linearly related.

In general, a focused image cannot be obtained from the standard two-dimensional motion compensation algorithm when three-dimensional target motion is present due to model mismatch. Therefore, it would be useful to detect the presence of three-dimensional motion directly from the radar data. Our approach is to utilize the AJTF algorithm to extract the phase behavior of the radar data at multiple range cells. We first parameterize the phase of the prominent point scatterer in one range cell using AJTF. Next we repeat the same procedure at another range cell. It can be shown that when the target undergoes only two-dimensional motion during the dwell duration, the ratio between the parameters ( $a_1, a_2, a_3$ ) extracted from one range cell and those corresponding parameters in another

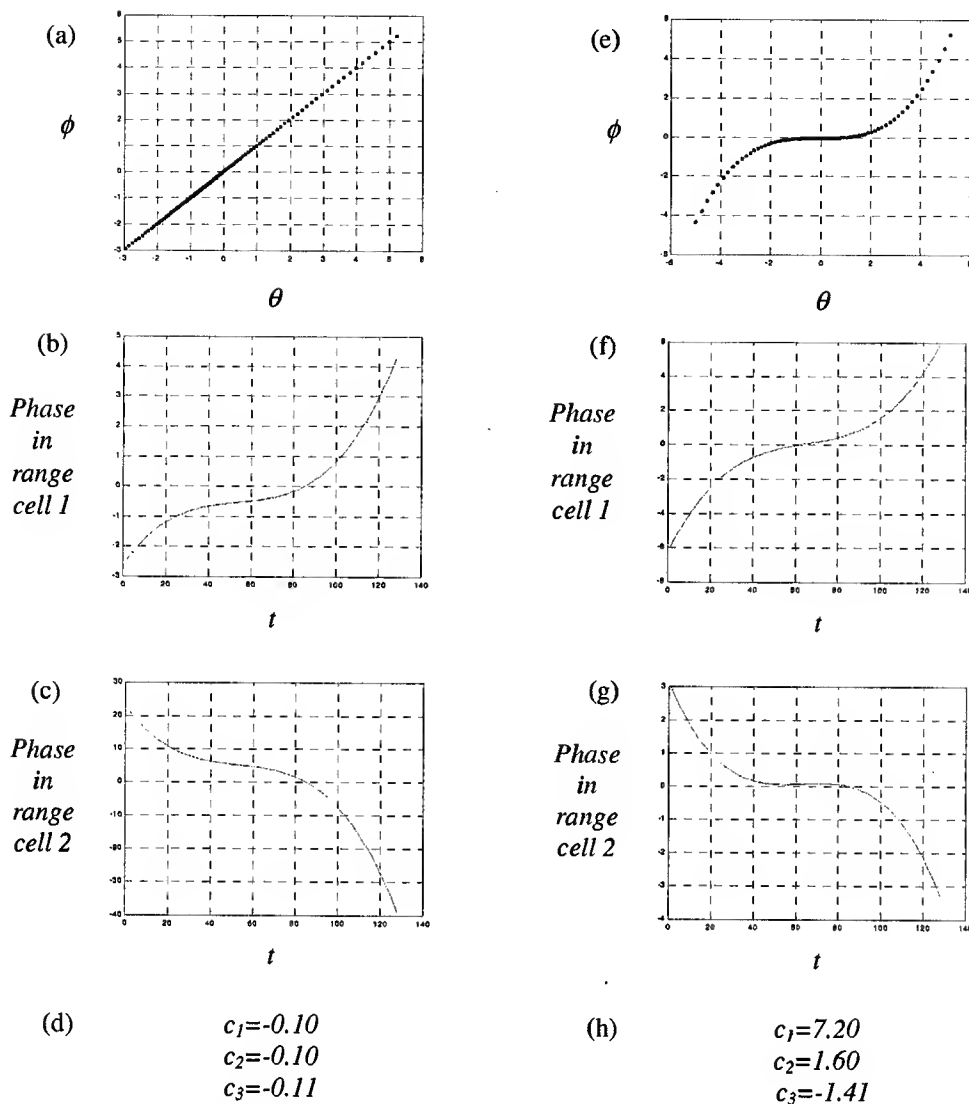


Fig. 3. (a) Simulated 2D target motion. (b) Phase behavior of the prominent point scatterer in range cell 1 extracted using AJTF. (c) Phase behavior of the prominent point scatterer in range cell 2 extracted using AJTF. (d) Ratios of the extracted phase parameters from the two range cells. Note that they are nearly constant. (e)-(h) Similar to (a)-(d), except that 3D motion is assumed. The resulting ratios in (h) are no longer constant.

range cell should be constant. Therefore, by examining the ratio of the parameters, we can distinguish two-dimensional motion from three-dimensional motion. Fig. 3 illustrates the idea using simulated point scatterer data. Figs. 3(a)-(d) show the two-dimensional motion scenario and Figs. 3(e)-(h) show the three-dimensional scenario. It can be seen from the results in Fig. 3(d) that the determined ratios:

$$c_i = a_i(\text{range cell 1}) / a_i(\text{range cell 2}) \quad (5)$$

are nearly constant for all the terms in case of two dimensional motion, as expected. For three-dimensional motion, the ratios are not the same, as seen in Fig. 3(h).

Fig. 4 shows our preliminary results of applying the 3D motion detection algorithm to real radar data. Fig. 4(a) shows the degree of three-dimensional motion in the data for 20 different image frames, detected by applying our algorithm to the raw radar data. As a reference for comparison, Fig. 4(b) shows the degree of three-dimensional motion for the same 20 frames measured using the motion data derived from inertial navigation instruments carried onboard the aircraft during data collection. It can be seen that our algorithm correctly detects where significant three-dimensional motions exist. We are currently fine tuning the algorithm to achieve faster and more robust detection. We believe this detection algorithm could be quite useful for determining the "good"

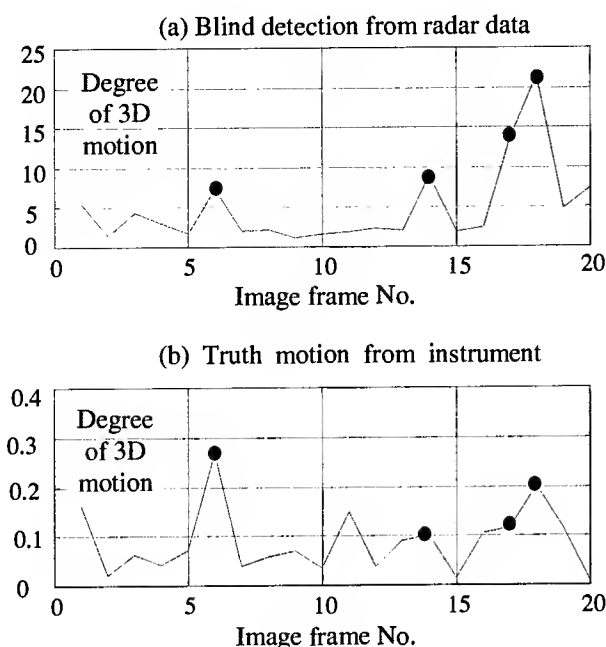


Fig. 4. Blind detection of three-dimensional motion from real radar data. (a) Degree of three-dimensional motion over 20 image frames detected using the proposed algorithm. (b) Degree of three-dimensional motion measured from on-board instrument data.

imaging intervals from which focused images can be more readily generated. For targets that exhibit very chaotic motions, such as ships on the ocean, finding such intervals of opportunity may be very critical for target recognition.

#### 4. SUMMARY

In this paper, we presented two applications of the adaptive joint time-frequency algorithm for ISAR image formation. In the first application, we carry out fine motion compensation to form focused ISAR images of articulating targets. The AJTF algorithm is used to estimate the phase of the prominent point scatterer within a range cell. The higher-order phase error due to uncompensated translation and rotational errors are then removed prior to the image formation. Results show that well-focused images can be obtained from measured data of an in-flight aircraft. In the second application, we try to detect the presence of three-dimensional target motion, for which a well-defined imaging plane does not exist. A three-dimensional motion model is utilized and the linearity of the phase functions of the prominent point scatterers between different range cells is used to distinguish two-dimensional from three-dimensional motion. The AJTF engine is again used to extract the phase function of the prominent scatterer within each range cell. Preliminary test results using real radar data indicate

that the algorithm can be used to detect those imaging intervals where conventional two-dimensional motion assumption would fail. We are working to devise algorithms for forming focused images even in the presence of these three-dimensional motions.

#### 5. ACKNOWLEDGMENT

This work is supported by the Office of Naval Research under Contract No. N00014-98-1-0615.

#### 6. REFERENCES

- [1] V. C. Chen, "Reconstruction of inverse synthetic aperture images using adaptive time-frequency wavelet transforms," *SPIE Proc. on Wavelet Application*, vol. 2491, pp. 373-386, 1995.
- [2] Y. Wang, H. Ling and V. C. Chen, "ISAR motion compensation via adaptive joint time-frequency technique," *IEEE Trans. Aerospace Electron. Syst.*, vol. 34, pp.670-677, Apr. 1998.
- [3] S. Qian and D. Chen, "Signal representation using adaptive normalized Gaussian functions," *Signal Processing*, vol. 36, no. 1, pp. 1-11, Mar. 1994.
- [4] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. Signal Processing*, vol. 41, pp. 3397-3415, Dec. 1993.
- [5] V.C. Chen and W. J. Miceli, "Effect of roll, pitch and yaw motions on ISAR imaging," *SPIE Proc. on Radar Processing*, vol. 3810, July 1999.
- [6] A. W. Rihaczek and S. J. Hershkowitz, "Choosing imaging intervals for identification of small ships," *SPIE Proc. on Radar Processing*, vol. 3810, July 1999.
- [7] J. Li, Y. Wang, R. Bhalla, H. Ling and V. C. Chen, "Comparison of high-resolution ISAR imagery from measured data and synthetic signatures," *SPIE Proc. on Radar Processing*, vol. 3810, July 1999.

# JOINT TIME-FREQUENCY ANALYSIS OF SAR DATA

*Ralph Fiedler and Robert Jansen*

NAVAL RESEARCH LABORATORY  
4555 OVERLOOK AVE SW  
WASHINGTON, D.C. 20375, USA

## ABSTRACT

The image formation process associated with coherent imaging sensors is particularly sensitive to and is often corrupted by non-stationary processes. In the case of synthetic aperture radar (SAR), non-stationary processes result from motion within the scene, variable radar cross section, multi-path, topographic variations, sensor anomalies, and deficiencies in the image formation processing chain. This paper addresses SAR image formation processing, the complex response function for a point source, and SAR JTF image formation implementations. Each of these topics is described within the context of applying JTF processing to all aspects of SAR image formation and analysis.

## 1. INTRODUCTION

The fundamental attribute of a synthetic aperture radar (SAR) is its ability to directly sample the complex Fourier domain of the spatial reflectivity map from an illuminated ground patch. This reflectivity map, or radar image, is the standard product from a SAR sensor. Apart from issues of sensor motion compensation and other corrections accounted for by the image formation processor the data collected by a SAR sensor is related to the desired radar image through a 2-dimensional Fourier transform. Key to the understanding and interpretation of SAR imagery is the realization that the received radar pulses are sampling the Fourier domain at different times and, when taken as a whole, fill a small annular region of the Fourier plane. The finite time scale of a SAR coherent data period (fraction of a minute) and the limited coverage in the Fourier plane (angular span of a few degrees and radial extent in proportion to the pulse fractional bandwidth) combine to form the root cause for the presence of and sensitivity to non-stationary processes in SAR data. Although non-stationary processes can degrade radar image quality and introduce peculiar signature artifacts, the sensitivity of SAR sensors to non-stationary processes provides an outstanding exploitation opportunity no incoherent imaging system can attest to. High-resolution SAR sensors are the best data sources for non-stationary signal exploitation since they span the longest coherent data period and have the largest range bandwidth.

Moreover, effective analysis of non-stationary processes can lead to their removal from the standard product yielding higher quality imagery.

The analysis of non-stationary processes in SAR data is necessarily a clutter, not a noise, dominated problem. Imaged scenes generally include some combination of urban infrastructure, vegetative ground cover, terrain features, water, and moving targets. Although these scene content categories contribute to stationary and non-stationary signal processes in SAR data, stationary processes tend to dominate most scenes. If this were not the case, the value of SAR imagery would be greatly diminished. Stationary processes are considered clutter within the context of non-stationary signal analysis.

The inevitable presence of non-stationary processes in SAR data spanning any real scene compels some form of JTF analysis. However the bi-linear character of traditional JTF analysis typically requires some form of filtering to mitigate the effects of the confusing cross terms, an overwhelming source of interference for a filled aperture SAR sensor. Here, 'filled aperture' refers to significant reflectivity over the entire imaging patch as opposed to the unfilled apertures of Inverse SAR (ISAR) imaging of ships and aircraft [1]. The preponderance of these interference terms limits broad utility of current JTF approaches within the context of SAR signal processing for single-phase center and single-frequency systems. Considerations of the underlying assumptions of SAR image formation processing together with the rich content of any real scene suggest a future developmental path comprising data driven JTF techniques focussing on the separability of stationary and non-stationary processes.

The exploitation of non-stationary processes in SAR data can be facilitated through joint time-frequency (JTF) signal processing. The most widely used JTF technique is the short-time Fourier transform (STFT). STFT processing in the parlance of SAR analysis is often referred to as sub-aperture processing. The SAR aperture that is synthesized over time by the relative motion between the sensor platform and the aim point is subdivided into smaller segments resulting in improved

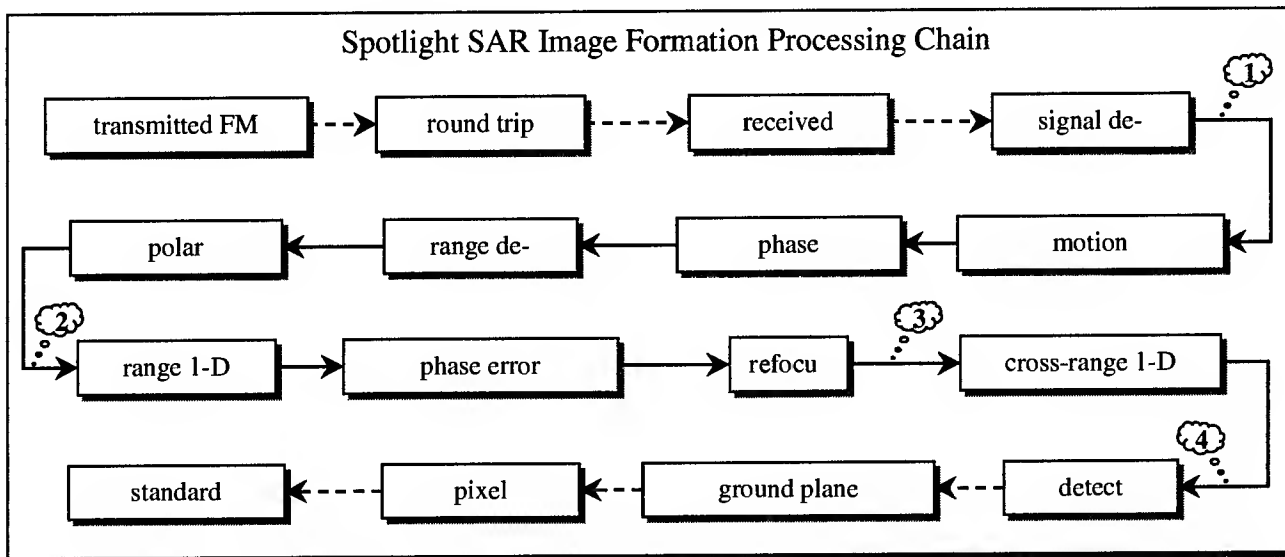


Figure 1. The SAR image formation processing chain may be generalized into the computational elements indicated in this flow diagram. The dashed arrows represent real signals and magnitude imagery, whereas the solid arrows represent complex signals and imagery. The callouts, or data stream taps, indicate locations in the processing chain where we consider joint time-frequency signal processing may be of benefit. See the text for a description of the taps.

presentation of non-stationary signatures, such as moving targets, but at the expense of degraded azimuthal resolution. A series of STFT sub-apertures created in this way form a three-dimensional data volume.

One advanced JTF technique relies on the Wigner-Ville distribution (WVD) characterizing both stationary and non-stationary processes without any degradation in resolution. The stationary and non-stationary components in the WVD time-frequency representation are self-terms of the underlying bi-linear distribution. The cross terms of the bi-linear distribution, however, introduce artifacts so severe as to render the utility of the JTF-WVD data volume unsuitable for many SAR related exploitation purposes. Each scattering center represented in the time-frequency plane mixes with every other scattering center regardless of whether the scattering center is stationary or not. If we take signatures represented in the time-frequency plane in pairs, the result of WVD is to introduce artificial signatures at a location half way between each pair with an amplitude greater than either signature taken individually. The superset of bi-linear JTF distributions is Cohen's class of distributions [2]. Various filtering mechanisms [3] have been developed in an attempt to reduce the cross term effects of WVD and the many other instances of Cohen's class of distributions.

With the exception of very low clutter environments, JTF analysis of SAR data should begin with the segmentation of stationary clutter from non-stationary signals. In support of eventual automated signal analysis, considerable importance should be placed on having the signal data drive the available degrees of freedom afforded by JTF algorithms. The principle degree of freedom, or adjustable parameter, is typically related to an area of regard such as the window width for the STFT.

## 2. SAR IMAGE FORMATION

The essential stages of SAR image formation processing are presented to illustrate the data stream taps where we consider JTF processing can benefit the analysis of SAR data. See Figure 1. Most of the discussion of SAR image formation in this section centers on spotlight-mode processing [4] [5]. Spotlight-mode SAR attains the best resolution of the possible collection modes of a SAR. During the coherent data period, the antenna is steered to remain pointed at a fixed aim point. The cross-range resolution is governed principally by the angular extent of the dwell on the aim point. Conversely, the antenna of a stripmap-mode SAR images broadside to the platform velocity vector resulting in a cross-range resolution equal to  $\frac{1}{2}$  the effective antenna diameter. In either case, the bandwidth of the pulse waveform is often chosen to

provide a range resolution comparable to that in the cross-range direction.

## 2.1 Processing Chain

The objectives for JTF SAR processing fall into two categories. One is value-added exploitation of the targets in the scene, and the other is enhancing or redefining the mechanics of the image formation process. Of the data stream taps indicated in Figure 1, tap {4} is the most readily available from commercial systems. Gaining access to the other taps normally requires direct access to an image formation processor, usually tailored to a specific system, so that the necessary modifications can be made. The JTF approaches we have under investigation are arranged by tap number.

1. This is the first useful tap in the SAR processing chain where the in-phase and quadrature-phase signals are formed into a complex data stream. JTF exploitation can potentially begin here on a pulse by pulse basis for problems having pulse width timescales. JTF signal processing may also benefit the phase correction stages leading up to polar formatting. The pulse data are represented here in a polar coordinate system. The polar formatting step resamples the pulse data from polar to Cartesian coordinate systems in preparation for 2-D Fast-Fourier transforms (FFT).
2. A 2-D complex Fourier domain map is rendered in the Cartesian coordinate system without any refocus applied. This tap is one starting point for higher dimensional JTF processing and exploitation. Performing JTF processing on each range line and then on each cross-range line will result with a 4-D data volume. Alternatively, applying a FFT along either dimension and then applying JTF processing to the other dimension results in a more manageable 3-D data volume. We refer to JTF processing along the cross-range direction as slow-time image formation processing (ST-IFP) processing, and along the range direction as fast-time image formation processing (FT-IFP) processing. JTF signal processing may also benefit the refocusing stages prior to the cross-range FFT.

3. Applying an inverse 1-D range FFT to this tap returns us to tap {2} with the added benefit of improved focus. Moreover, since range compression has already occurred, ST-JTF processing can proceed directly from this tap. For these reasons we consider tap {3} to be of greater practical use than tap {2} for most problems.
4. This tap provides the slant plane complex image. The slant plane is the plane formed by the platform velocity vector and the range line to the aim point. From this point, JTF processing can proceed after an inverse 1-D range and/or 1-D cross-range FFT is performed. For general JTF exploitation, this tap is the most convenient. Earlier taps are required if JTF enhancements to the image formation process are to be explored.

The annulus sampled by a SAR in the complex Fourier domain is defined by the angle subtended by the dwell of the sensor on the aim point and by the bandwidth of the pulse waveform. Even if there were no motion in the scene non-stationary processes can still be expected. Coherent response from structures comprising linear, planar, dihedral, and trihedral elements result in correlated phase in the complex image domain. As a result, the computed reflectivity map will vary between selected annuli in the Fourier domain. Equivalently, the reflectivity of man-made structures is aspect dependent. Conversely, a scene dominated by random scattering processes will result with a reflectivity map that is independent of the subset selected, apart from differences in the speckle content.

Although many of the stages depicted in the processing flow diagram are designed to correct for platform motion and sampling artifacts, the range de-skew stage is directly related to the formulation of the SAR response function. Range skew is a phase term that represents a departure of the SAR response function from the 2-D Fourier transform of the desired radar reflectivity map. The de-skew correction is therefore accomplished prior to the 1-D FFT, or compression, stages. The occurrence of range dependent skew in the SAR response function is highlighted in the next section.

## 2.2 Normalized Response Function

The normalized SAR response function for a point source with complex reflectivity  $g_0$  may be expressed as

$$\hat{r}_\theta(\hat{t}) = \frac{1}{2} A_\theta(\vec{p}) g_\theta e^{2\pi i \left( 2 \frac{\epsilon}{Q} \mu_\theta^2 - 2 \left( 1 + \frac{\tau}{Q} \right) \mu_\theta \right)}, \quad (1)$$

where  $A_\theta(\vec{p})$  encompasses complex antenna gain and any propagative effects,  $\vec{p}$  is the position vector of the point source relative to the ground reference point,  $Q$  is the ratio of the center frequency to the chirp waveform bandwidth,  $\tau$  is the pulse time normalized by the pulse duration,  $\mu$  is the range offset to the point source relative to the ground reference point and is normalized by the center frequency wavelength,  $\theta$  is the polar angle described by the motion of the platform as seen by the aim point, and  $\epsilon$  is the inverse of the mean number of cycles transmitted during the duration of a radar pulse. Although the frequency and bandwidth of a SAR define the character and resolution of a radar image, the system  $Q$  is key to the dynamic range available to the sensor. The smaller the  $Q$ , or the larger the fractional bandwidth, the more robust is the sampling of the complex phase domain. The derivation of (1) extends from pioneering tomographic approaches to spotlight-mode SAR processing [6].

The first phase term in (1) causes a range dependent distortion, or image skew. This term is removed from the SAR signal history by the range de-skew stage illustrated in Figure 1. Although the magnitude of this phase term varies as the square of the range offset from the aim point, the coefficient  $\epsilon$  is sufficiently small to allow this term to be neglected in many cases.

After the image de-skew corrections are applied to the signal history, the SAR response function is seen to reduce to the Fourier transform of a point source projected along a polar angle  $\theta$ . Pulse data collected over a sufficiently large polar annulus can then be resampled to a Cartesian grid and inverse Fourier transformed to produce the radar image. This approach to SAR image formation is accurate if there are only stationary processes present in the signal history. An examination of (1) shows many sources where non-stationary processes can be introduced.

- $A_\theta(\vec{p})$  - (a) The antenna beam pattern and the image patch size are chosen to minimize image quality degradation. Beam de-shading performed after the detection stage in Figure 1 will correct for beam related intensity rolloffs. Apart from sensor hardware instabilities that affect the complex antenna gain, antenna properties

are not considered to be a significant contributor to non-stationary processes. (b) Although radar is generally considered a day, night, and all weather sensor, electrical storms can introduce propagative anomalies that will affect image quality. Spaceborne SARs may further be affected by inhomogeneities and fluctuations in the electron density of the ionosphere.

- $g_\theta$  - The complex reflectivity of a point source can vary over the polar angle spanned by the coherent data period. This is especially true for linear, planar, and dihedral structures whose principle attribute for non-stationary processes is that they have very narrow beam patterns.
- $\mu_\theta$  - Time dependent variations in the range offset due to motion in the scene is the most popular issue addressed by researchers exploring JTF applications for imaging radars. Mover defocus resulting from range acceleration, cross-range velocity, and cross-range acceleration can be enhanced using JTF techniques, not only for just one mover, but simultaneously for all movers in the scene. Whereas, JTF techniques may be effective for the sparse scenes of inverse synthetic aperture radar (ISAR), e.g., the imaging of ships or planes, the stationary clutter dominated scenes of SAR introduce an overwhelming source of cross-terms in bi-linear JTF techniques that make it difficult to effectively analyze embedded non-stationary processes. Techniques for filtering out stationary clutter are needed to exploit non-stationary signals beyond the fidelity available from traditional STFT approaches.

### 3. CONCLUSIONS

The apparent utility of JTF techniques for SAR data analysis is significantly affected by cross-terms associated with the bi-linear distributions commonly employed in the field of JTF signal processing. SAR data are typically dominated by the clutter of stationary processes, e.g., urban infrastructure, vegetation, and natural terrain. Mixed within that clutter are non-stationary signals. Cross terms in the time-frequency domain therefore arise from clutter-to-clutter mixing and clutter to non-stationary signals mixing. Despite

the many techniques developed to mitigate the effects of the cross-terms, the dominance of stationary clutter in most SAR data is overwhelming. Research into the separability of non-stationary signals from stationary clutter coupled with signal-based, or adaptive, JTF techniques appears to be most promising approach for extending JTF signal processing of SAR exploitation beyond STFT techniques.

## REFERENCES

- [1] V. C. Chen and H. Ling, "Joint Time-Frequency Analysis for Radar Signal and Image Processing", *IEEE Signal Processing Magazine*, 81-93, March 1999.
- [2] L. Cohen, *Time-Frequency Analysis*, Prentice-Hall PTR, Upper Saddle River, 1995.
- [3] S. Qian and D. Chen, *Joint Time-Frequency Analysis: Methods and Applications*, Prentice-Hall PTR, Upper Saddle River, 1996.
- [4] C. V. Jakowatz, Jr., D. E. Wahl, P. H. Eichel, D. C. Ghiglia, and P. A. Thompson, *Spotlight-Mode Synthetic Aperture Radar: A Signal Processing Approach*, Kluwer Academic Publishers, Boston, 1996.
- [5] W. G. Carrara, R. S. Goodman, and R. M. Majewski, *Spotlight Synthetic Aperture Radar: Signal Processing Algorithms*, Artech House, Boston, 1995.
- [6] D. C. Munson, Jr., J. D. O'Brien, and W. K. Jenkins, "A Tomographic Formulation of Spotlight-Mode Synthetic Aperture Radar", *Proceedings of the IEEE*, **71**, 917-925.



# PULSE PROPAGATION IN DISPERSIVE MEDIA

Leon Cohen

City University of New York, 695 Park Ave., New York, NY 10021 USA.

## ABSTRACT

We give a simple formula for the calculation of the moments of a propagating pulse in a dispersive medium. Both the spatial and time moments are considered. Explicit formulas are derived for the spatial spreading of a propagating pulse and also for the duration of the pulse at a fixed position in space. In addition, we give formulas for the calculation of the instantaneous frequency of a pulse at a given position. A number of simple examples are used to illustrate the formulas derived.

## 1. INTRODUCTION

Linear partial differential equations whose solutions give wave like behavior come in many varieties, but fortunately, the solution to all of them can be written in a simple form [4,5]. We call the solution to such an equation  $u(x, t)$  where  $x$  and  $t$  are the spatial and time variable respectively. A general method of solution is to substitute

$$e^{ikx - i\omega t} \quad (1)$$

into the wave equation with the result that such a particular solution can only exist if there is a relationship between  $k$  and  $\omega$ . The relationship will be of the form

$$D(\omega, k) = 0 \quad (2)$$

This is called dispersion relation. One can now solve for  $k$  in terms of  $\omega$  or the other way around. These relations are written here as

$$k = K(\omega) \quad ; \quad \omega = W(k) \quad (3)$$

Generally there will be more than one solution and each solution is called a mode. Furthermore depending on whether we have complex or real solutions we will have damping or not. Here, we consider the case where we have no damping, that is both  $K(\omega)$  and  $W(k)$  are real.

The general solution for  $u(x, t)$  is then expressed in terms of Fourier integrals taking into account the

Work supported by the Office of Naval Research, the NASA JOVE, and the NSA HBCU/MI programs.

dispersion relation. This is described in Sections 2 and 3. However, there are two distinct physical situations depending on the initial conditions. The two types of initial conditions are

$$\text{Given: } u(x, 0) \quad (\text{Case One}) \quad (4)$$

$$\text{Given: } u(0, t) \quad (\text{Case Two}) \quad (5)$$

The first case is when we have the spatial wave at a given time and the second when we have the wave at a given position for all time. An example of the first is if we pluck a string and let go at time zero. An example of the second is if we are at a fixed position and create a pulse, for example, a radar, sonar, or fiber optic pulse.

### Group Velocity and Its Extension

A central idea in the study of pulse propagation is the group velocity,  $v_g(k)$ , which is given by

$$v_g(k) = \omega'(k) \quad (6)$$

There are many plausible arguments that have been given in the literature for calling this quantity the group velocity. In Sec. 2 we will give a new relation for a propagating pulse that we think gives a very clear picture why  $v_g(k)$  should be called a group velocity and how it is related to the propagation of the center of mass of the pulse.

In Sec. 3 we will study the time properties of a pulse at a fixed position. We will see that the natural quantity that appears is

$$z_g(\omega) = K'(\omega) \quad (7)$$

We note that it has the units of inverse velocity. We will see that it is related to the amount of time delay per unit distance. We shall call it the group time delay.<sup>1</sup>

### Instantaneous Frequency

A pulse can always be written in terms of its amplitude and phase

$$u(x, t) = |u(x, t)|e^{i\varphi_a(x, t)} \quad (8)$$

<sup>1</sup>This quantity should not be confused with "group delay", which is the derivative of the spectral phase [3].

The instantaneous frequency at a fixed position is given by the partial derivative of the phase with respect to time <sup>2</sup>

$$\omega_i(x, t) = \pm \frac{\partial}{\partial t} \varphi(x, t) \quad (9)$$

## 2. CASE ONE

*General Solution.* We now consider the situation where the initial condition is given by Eq. (4). The general solution is [4,5]

$$u(x, t) = \frac{1}{\sqrt{2\pi}} \int S(k) e^{ikx - iW(k)t} dk \quad (10)$$

where  $S(k)$  is the initial spatial spectrum

$$S(k) = \frac{1}{\sqrt{2\pi}} \int u(x, 0) e^{-ikx} dx \quad (11)$$

We define the time dependent spectrum by

$$S(k, t) = S(k, 0) e^{-iW(k)t} \quad (12)$$

where

$$S(k, 0) = S(k) \quad (13)$$

Therefore

$$u(x, t) = \frac{1}{\sqrt{2\pi}} \int S(k, t) e^{ikx} dk \quad (14)$$

$$S(k, t) = \frac{1}{\sqrt{2\pi}} \int u(x, t) e^{-ikx} dx \quad (15)$$

and hence  $u(x, t)$  and  $S(k, t)$  form Fourier transform pairs for all time. Since  $u(x, t)$  and  $S(k, t)$  form Fourier transform pairs we can use the operator method to calculate moments [1,2,3]

$$(x^n)_t = \int x^n |u(x, t)|^2 dx \quad (16)$$

$$= \int S^*(k, t) \mathcal{X}^n S(k, t) dk \quad (17)$$

where  $\mathcal{X}$  is the position operator in the  $k$  representation

$$\mathcal{X} = i \frac{\partial}{\partial k} \quad (18)$$

*Moments.* Defining the group velocity,  $v_g(k)$ , by,

$$v_g(k) = W'(k) \quad (19)$$

<sup>2</sup>Whether one takes  $\pm$  in Eq. (9) depends on the form taken for Eq. (1). In particular,  $\pm$  should be chosen to be the same as the sign in front of  $\omega t$  in Eq. (1). For the choice taken here the negative sign should be used in Eq. (9).

the *exact* first two moments and standard deviation are worked out to be [1,2]

$$(x)_t = (x)_0 + Vt \quad (20)$$

$$(x^2)_t = (x^2)_0 + t^2 (v_g^2)_0 + t(v_g \mathcal{X} + \mathcal{X} v_g)_0 \quad (21)$$

$$\sigma_{x|t}^2 = \sigma_{x|0}^2 + 2t \text{Cov}_{xv_g} + t^2 \sigma_{v_g}^2 \quad (22)$$

where

$$V = \int v_g(k) |S(k, 0)|^2 dk \quad (23)$$

$$\sigma_{v_g}^2 = \int (v_g(k) - V)^2 |S(k, 0)|^2 dk \quad (24)$$

$$\text{Cov}_{xv_g} = \frac{1}{2} (v_g \mathcal{X} + \mathcal{X} v_g)_0 - (v_g)_0 (x)_0 \quad (25)$$

An alternative way to calculate the covariance is to first write  $S(k, 0)$  in terms of its amplitude and phase

$$S(k, 0) = |S(k, 0)| e^{i\psi(k, 0)} \quad (26)$$

It can be shown that [3]

$$\frac{1}{2} (v_g \mathcal{X} + \mathcal{X} v_g)_0 = - \int v_g(k) \psi'(k, 0) |S(k, 0)|^2 dk \quad (27)$$

*Asymptotic Solution.* The standard method to study Eq. (14) is the asymptotic solution which is obtained by the method of stationary phase [5]. The basic idea is to find the value of  $k$  where the contribution of the integrand is largest. The value of  $k$  is obtained from solving the equation [5]

$$W'(k) = x/t \quad (28)$$

for  $k$ . Then,

$$u_a(x, t) \sim S(k) \sqrt{\frac{2\pi}{tW''(k)}} e^{ikx - iW(k)t - i\pi \text{sgn } W''/4} \quad (29)$$

The amplitude and phase are

$$|u_a(x, t)| = |S(k)| \sqrt{\frac{2\pi}{tW''(k)}} \quad (30)$$

$$\varphi_a(x, t) = \psi(k) + kx - W(k)t - \pi \text{sgn } W''/4 \quad (31)$$

*Instantaneous Frequency.* Differentiating the phase,  $\varphi_a(x, t)$ , as given by Eq. (31) we have

$$\omega_i(x, t) = - \left[ \frac{d\psi}{dk} + x - t \frac{dW(k)}{dk} \right] \frac{\partial k}{\partial t} + W(k) \quad (32)$$

But by Eq. (28)

$$x - tW'(k) = 0 \quad (33)$$

and therefore

$$\omega_i(x, t) = -\frac{d\psi}{dk} \frac{\partial k}{\partial t} + W(k) \quad (34)$$

Also, from Eq. (28) we have

$$W''(k) \frac{\partial k}{\partial t} = -x/t^2 \quad (35)$$

giving

$$\frac{\partial k}{\partial t} = -\frac{x}{t^2 W''(k)} = -\frac{W'(k)}{t W''(k)} = -\frac{W^2(k)}{x W''(k)} \quad (36)$$

Any one of these can be substituted into Eq. (34) to obtain

$$\omega_i(x, t) = \frac{x}{t^2 W''(k)} \frac{d\psi}{dk} + W(k) \quad (37)$$

$$= \frac{W'(k)}{t W''(k)} \frac{d\psi}{dk} + W(k) \quad (38)$$

$$= \frac{W'^2(k)}{x W''(k)} \frac{d\psi}{dk} + W(k) \quad (39)$$

### 3. CASE TWO

*General Solution.* The general solution is

$$u(x, t) = \frac{1}{\sqrt{2\pi}} \int F(\omega) e^{iK(\omega)x - i\omega t} d\omega \quad (40)$$

where  $F(k)$  is the initial time spectrum at  $x = 0$ ,

$$F(\omega) = \frac{1}{\sqrt{2\pi}} \int u(0, t) e^{i\omega t} dt \quad (41)$$

We point out that  $u(0, t)$  is what is usually called a "signal". We define the space dependent spectrum by

$$F(\omega, x) = F(\omega, 0) e^{iK(\omega)x} \quad (42)$$

where

$$F(\omega, 0) = F(\omega) \quad (43)$$

Hence,

$$u(x, t) = \frac{1}{\sqrt{2\pi}} \int F(\omega, x) e^{-i\omega t} d\omega \quad (44)$$

$$F(\omega, x) = \frac{1}{\sqrt{2\pi}} \int u(x, t) e^{i\omega t} dt \quad (45)$$

which shows that  $u(x, t)$  and  $F(\omega, x)$  form Fourier transform pairs for any  $x$ .

*Moments.* As before, we can write the time moments as

$$\langle t^n \rangle_x = \int t^n |u(x, t)|^2 dt \quad (46)$$

$$= \int F^*(\omega, x) (-T)^n F(\omega, x) d\omega \quad (47)$$

where  $T$  is the time operator in the frequency domain<sup>3</sup>,

$$T = i \frac{\partial}{\partial \omega} \quad (48)$$

Defining the group time delay,  $z_g(\omega)$ , by

$$z_g(\omega) = K'(\omega) \quad (49)$$

the exact first two moments and standard deviation are

$$\langle t \rangle_x = \langle t \rangle_0 + Zx \quad (50)$$

$$\langle t^2 \rangle_x = \langle t^2 \rangle_0 + x^2 \langle z_g^2 \rangle_0 + x \langle z_g T + T z_g \rangle_0 \quad (51)$$

$$\sigma_{t|x}^2 = \sigma_{t|0}^2 + 2x \text{Cov}_{tz_g} + x^2 \sigma_{z_g}^2 \quad (52)$$

where

$$Z = \int z_g(\omega) |F(\omega, 0)|^2 d\omega \quad (53)$$

$$\sigma_{z_g}^2 = \int (z_g(\omega) - Z)^2 |F(\omega, 0)|^2 d\omega \quad (54)$$

$$\text{Cov}_{tz_g} = \frac{1}{2} \langle z_g T + T z_g \rangle_0 - \langle z_g \rangle_0 \langle t \rangle_0 \quad (55)$$

Also, if we write

$$F(\omega, 0) = |F(\omega, 0)| e^{i\eta(\omega, 0)} \quad (56)$$

then

$$\frac{1}{2} \langle z_g T + T z_g \rangle_0 = - \int z_g(\omega) \psi'(\omega, 0) |F(\omega, 0)|^2 d\omega \quad (57)$$

We point out that  $\sigma_{t|x}$  is what is commonly called the duration of a signal. In this case it is a duration of the signal at position  $x$ . We see that for  $x \rightarrow \infty$  the duration must go to infinity no matter what the duration is at the point where it is generated.

*Asymptotic solution.* One obtains  $\omega$  from

$$K'(\omega) = t/x \quad (58)$$

and the asymptotic approximation is then

$$u_a(x, t) \sim F(\omega) \sqrt{\frac{2\pi}{x K''(\omega)}} e^{iK(\omega)x - i\omega t - i\pi \text{sgn } K''/4} \quad (59)$$

<sup>3</sup>The reason for the negative sign in  $(-T)^n$  is because of the way the Fourier transform was defined in Eq. (41).

$$|u_a(x, t)| = |F(\omega)| \sqrt{\frac{2\pi}{xK''(\omega)}} \quad (60)$$

$$\varphi_a(x, t) = \eta(\omega) + K(\omega)x - \omega t - \pi \operatorname{sgn} K''/4 \quad (61)$$

*Instantaneous Frequency.* Differentiating the phase we have

$$\omega_i(x, t) = -\frac{\partial}{\partial t} \varphi_a(x, t) = -\frac{\partial \omega}{\partial t} \left[ \frac{d\eta}{d\omega} + x \frac{dK(\omega)}{d\omega} - t \right] + \omega \quad (62)$$

ad using Eq. (58) we have that

$$\omega_i(x, t) = \left[ \frac{\partial \omega}{\partial t} \frac{d\eta}{d\omega} \right] + \omega \quad (63)$$

Also,

$$K''(\omega) \frac{\partial \omega}{\partial t} = 1/x \quad (64)$$

giving

$$\frac{\partial \omega}{\partial t} = \frac{1}{xK''(\omega)} = \frac{K'(k)}{tK''(k)} \quad (65)$$

Hence,

$$\omega_i(x, t) = \frac{1}{xK''(\omega)} \frac{d\eta}{d\omega} + \omega \quad (66)$$

$$= \frac{K'(k)}{tK''(k)} \frac{d\eta}{d\omega} + \omega \quad (67)$$

*Real Spectrum.* If the initial spectrum is real then  $\eta = 0$  and we have that

$$\omega_i(x, t) = \omega \quad ; \quad K'(\omega) = t/x \quad (68)$$

#### 4. EXACTLY SOLVABLE EXAMPLES

In all the examples we consider the case where the dispersion relation is given by

$$K(\omega) = \gamma \omega^2 \quad (69)$$

*Example 1.* We generate a pulse at  $x = 0$  which is a pure sinusoid

$$u(0, t) = e^{-j\omega_0 t} \quad (70)$$

Its spectrum is

$$F(\omega) = \sqrt{2\pi} \delta(\omega - \omega_0) \quad (71)$$

Putting this into Eq. (40) we obtain

$$e^{i\gamma\omega_0^2 x - j\omega_0 t} \quad (72)$$

and we see that the phase is given by

$$\varphi(x, t) = \gamma\omega_0^2 x + \omega_0 t \quad (73)$$

which gives

$$\omega_i = \omega_0 \quad (74)$$

for the instantaneous frequency. It is independent of the dispersion or position.

*Example 2.* Suppose we take an impulse at  $x = 0$

$$u(0, t) = \delta(t - t_0) \quad (75)$$

which gives

$$F(\omega) = \frac{1}{\sqrt{2\pi}} e^{i\omega t_0} \quad (76)$$

Using Eq. (40) we obtain

$$u(x, t) = \frac{1}{2\pi} \frac{i\pi}{\gamma x} e^{i\omega t_0} \exp \left[ -i \frac{(t - t_0)^2}{4\gamma x} \right] \quad (77)$$

The instantaneous frequency is

$$\omega_i(x, t) = \frac{t - t_0}{2\gamma x} \quad (78)$$

which is chirp.

*Example 3.* Consider the signal

$$u(0, t) = (\alpha/\pi)^{1/4} e^{-\alpha t^2/2 - j\omega_0 t} \quad (79)$$

whose spectrum is

$$F(\omega) = \frac{(\alpha/\pi)^{1/4}}{\sqrt{\alpha}} \exp \left[ -\frac{(\omega - \omega_0)^2}{2\alpha} \right] \quad (80)$$

Working out the solution we obtain

$$u(x, t) = \frac{(\alpha/\pi)^{1/4}}{\sqrt{2\alpha}} \sqrt{\frac{1}{\frac{1}{2}\alpha - i\gamma x}} \times \quad (81)$$

$$\exp \left[ -\frac{\omega_0^2}{2\alpha} + \frac{(it + \frac{\omega_0}{\alpha})}{4(\frac{1}{2}\alpha - i\gamma x)} \right] \quad (82)$$

The phase and amplitude are given by

$$\begin{aligned}\varphi(x, t) &= \frac{-\gamma x t^2 - \omega_0 t / \alpha^2 + \omega_0 \gamma x / \alpha^2}{4 \left[ \left( \frac{1}{2\alpha} \right)^2 + \gamma^2 x^2 \right]} + \frac{1}{2} \arctan \frac{2\gamma x}{\alpha} \\ &= \frac{-\gamma x \alpha^2 t^2 - \omega_0 t + \omega_0 \gamma x}{1 + 4\alpha^2 \gamma^2 x^2} + \frac{1}{2} \arctan \frac{2\gamma x}{\alpha}\end{aligned}\quad (83)$$

$$\begin{aligned}|u(x, t)| &= \frac{(\alpha/\pi)^{1/4}}{\sqrt{2\alpha}} \left( \frac{1}{\alpha/4 + \gamma^2 x^2} \right)^2 \times \\ &\exp \left[ -\frac{1}{2} \alpha \left( \frac{t^2 - 4\omega_0 \gamma x (t - \omega_0 \gamma x)}{1 + 4\alpha^2 \gamma^2 x^2} \right) \right]\end{aligned}\quad (84)$$

For the exact instantaneous frequency we have

$$\omega_i(x, t) = \frac{\omega_0 + 2\alpha^2 \gamma x t}{1 + 4\alpha^2 \gamma^2 x^2} \quad (85)$$

This is a chirp even though a pure sine wave is being generated at  $x = 0$ . In fact, even for  $\omega = 0$  we have a chirp.

*Example 4.* Consider

$$u(0, t) = e^{-i\beta t^2/2 - j\omega_0 t} \quad (86)$$

whose spectrum is

$$F(\omega) = \sqrt{\frac{1}{i\beta}} \exp \left[ -i \frac{(\omega - \omega_0)^2}{2\beta} \right] \quad (87)$$

The solution is

$$\begin{aligned}u(x, t) &= \sqrt{\frac{1}{1 - 2\gamma\beta x}} \times \\ &\exp \left[ -i \frac{\beta t^2/2 + \omega_0 t + \gamma \omega_0^2 x}{1 - 2\gamma\beta x} \right]\end{aligned}\quad (88)$$

For the exact instantaneous frequency we have

$$\omega_i(x, t) = \frac{\beta t + \omega_0}{1 - 2\gamma\beta x} \quad (89)$$

Therefore, we still have a chirp but the chirp rate changes with distance.

*Example 5.* Now consider the asymptotic solution for Example 4. Solving for  $\omega$  from

$$K'(\omega) = 2\gamma\omega = \frac{t}{x} \quad (90)$$

we have

$$\omega = \frac{t}{2\gamma x} \quad (91)$$

Therefore,

$$F\left(\omega = \frac{t}{2\gamma x}\right) = \frac{(\alpha/\pi)^{1/4}}{\sqrt{\alpha}} \exp \left[ -\frac{(t - 2\gamma\omega_0 x)^2}{8\gamma^2 x^2 \alpha} \right] \quad (92)$$

Using Eq. (59) we have that

$$\begin{aligned}u(x, t) &= \frac{(\alpha/\pi)^{1/4}}{\sqrt{2\alpha}} \sqrt{\frac{\pi}{\gamma x}} \times \\ &\exp \left[ -\frac{(t - 2\gamma\omega_0 x)^2}{8\gamma^2 x^2 \alpha} - i \frac{t^2}{4\gamma x} - i\pi \operatorname{sgn} \gamma / 4 \right]\end{aligned}\quad (93)$$

This gives an instantaneous frequency given by

$$\omega_i = \frac{t}{2\gamma x} \quad (94)$$

The instantaneous frequency could be obtained directly from Eq. (66). For this case we have that  $\eta = 0$  and hence

$$\omega_i = \omega = \frac{t}{2\gamma x} \quad (95)$$

## 5. CONCLUSION

We have given simple formulas for the moments, spread, and instantaneous frequency, of a propagating pulse in dispersive media.

## REFERENCES

- [1] L. Cohen, "Characterizations of Transients", *SPIE Proceedings*, vol. 3069, pp. 2-15, 1997.
- [2] L. Cohen, "Series Expansion for a Propagating Pulse", *NCA-Vol.24*, 77-82, 1997.
- [3] L. Cohen, *Time-Frequency Analysis*, Prentice-Hall, Englewood Cliffs, 1995.
- [4] A. J. D. Jackson, *Classical Electrodynamics*, Wiley, 1992.
- [5] G. B. Whitham, *Linear and Nonlinear Waves*, Wiley, 1974.

# WAVELET-BASED MODELS FOR NETWORK TRAFFIC

Dong Wei and Haiguang Cheng

Center for Telecommunications and Information Networking  
Department of Electrical and Computer Engineering, Drexel University  
Philadelphia, PA 19104 U.S.A.  
E-mails: wei@ece.drexel.edu, hgcheng@io.ece.drexel.edu

## ABSTRACT

We propose a novel set of wavelet-based stochastic models for self-similar network traffic with non-Gaussian behaviors. We show that these models are sufficiently accurate and parsimonious, and have very low computational complexity in analysis and synthesis.

## 1. INTRODUCTION

Recent studies of high-quality, high-resolution network traffic measurements have revealed that packet traffic appears to be both self-similar (or long-range dependent) and non-Gaussian distributed [1]. In order to realize the desirable properties of communication networks, such as ubiquity, convenience, affordability, reliability, and security, it is crucial to develop accurate and efficient traffic models that are capable of yielding acceptably precise performance predictions in a reasonable amount of time.

A real-valued stochastic process  $X(t)$  is said to be *statistically self-similar* with parameter  $H$  if for any  $a > 0$ ,

$$X(t) = a^{-H} X(at) \quad (1)$$

where the equality holds in a statistical sense (e.g., in all finite-dimensional joint distributions or in second-order statistics) and  $H$  is the so-called Hurst parameter, which satisfies  $0 < H < 1$  and captures the degree of self-similarity. In the context of network traffic, which are typically modeled as non-negative processes, the Hurst parameter can be used as a measure of burstiness [2]. The time-averaged spectrum of  $X(t)$ , denoted by  $S_X(\omega)$ , exhibits a  $1/f$  behavior:

$$\frac{C_l}{|\omega|^{2H-1}} \leq S_X(\omega) \leq \frac{C_u}{|\omega|^{2H-1}} \quad (2)$$

where  $C_l$  and  $C_u$  are constants satisfying  $0 < C_l \leq C_u < \infty$ .

This work was supported by Defense Advanced Research Project Agency under grant F30602-00-2-0501.

The multiscale property of wavelets [3], [4] makes wavelet representations to be natural and powerful analysis and synthesis tools for self-similar network traffic. A wavelet-domain independent Gaussian model is proposed in [5]. A wavelet-based multi-fractal model is proposed in [6].

In this paper, we propose a novel set of wavelet-based stochastic models for the emerging complex high-speed packet network traffic with self-similar and non-Gaussian behaviors. We show that these models are sufficiently accurate and parsimonious, and have very low computational complexity in analysis and synthesis.

The following convention of notation is used in the paper:

$$\sum_k \equiv \sum_{k=-\infty}^{\infty} \quad (3)$$

## 2. WAVELET-BASED SYNTHESIS OF NON-NEGATIVE SELF-SIMILAR PROCESSES

### 2.1. Wavelet-Based Models

Let  $\psi$  and  $\phi$  be the synthesis wavelet and scaling function of a two-channel, compactly supported, real-valued, biorthogonal wavelet system [3], respectively. We construct a random process  $X(t)$  by means of a biased wavelet series expansion:

$$X(t) = \sum_i \sum_k W_{i,k} \psi_{i,k}(t) + \mu \quad (4)$$

where we have used the short-hand notation

$$\psi_{i,k}(t) = 2^{i/2} \psi(2^i t - k) \quad (5)$$

for the dilated and translated versions of  $\psi(t)$ , and hereafter we shall apply the notation to  $\phi$  similarly. The wavelet coefficients  $\{W_{i,k} : k \in \mathbf{Z}\}$  at scale  $2^i$

are independent, identically distributed (i.i.d.) random variables with zero mean and variance

$$E[W_{i,k}^2] = \sigma_i^2 = 2^{-i(2H-1)} \sigma^2 \quad (6)$$

where  $\sigma^2$  is a reference variance. The constant  $\mu$  is used to represent the desired mean of the process. According to [4, Theorem 3.4],  $X(t)$  is a  $1/f$  process.

However, it is impractical to synthesize the random process  $X(t)$  using (4) due to that infinitely many scales are required. Therefore, the method suggested by the theorem is not useful in practice.

We propose to synthesize a process  $X_I(t)$  using a finite number of scales in the wavelet series expansion:

$$X_I(t) = \sum_k S_{I,k} \phi_{I,k}(t) + \mu \quad (7)$$

$$= \sum_k S_{i_0,k} \phi_{i_0,k}(t) + \sum_{i=i_0}^{I-1} \sum_k W_{i,k} \psi_{i,k}(t) + \mu \quad (8)$$

$$= \sum_{i=-\infty}^{I-1} \sum_k W_{i,k} \psi_{i,k}(t) + \mu \quad (9)$$

where  $\{S_{i,k} : k \in \mathbf{Z}\}$  are the unbiased scaling coefficients at scale  $2^i$ . In order to obtain non-negative processes, we choose to use biorthogonal  $B$ -spline wavelets [3] whose synthesis scaling functions are non-negative. By comparing (4) and (9), we obtain

$$\lim_{I \rightarrow \infty} X_I(t) = X(t) \quad (10)$$

which implies that  $X_I(t)$  is an asymptotically  $1/f$  process.

We use an iterative procedure to synthesize  $X_I(t)$  according to (8):

Step 1: set  $i := i_0$ ;

Step 2: synthesize  $\{S_i[k] : k \in \mathbf{Z}\}$ , the scaling coefficients at scale  $2^i$ ;

Step 3: synthesize  $\{W_i[k] : k \in \mathbf{Z}\}$ , the wavelet coefficients at scale  $2^i$ , from  $\{S_i[k] : k \in \mathbf{Z}\}$ ;

Step 4: synthesize  $\{S_{i+1}[k] : k \in \mathbf{Z}\}$ , the scaling coefficients at scale  $2^{i+1}$ , from  $\{S_i[k] : k \in \mathbf{Z}\}$  and  $\{W_i[k] : k \in \mathbf{Z}\}$  by means of the Mallat synthesis algorithm [7]:

$$S_{i+1,k} = \sum_l (h[k-2l] S_{i,l} + g[k-2l] W_{i,l}) \quad (11)$$

where  $h[k]$  and  $g[k]$  are the FIR synthesis filters of the wavelet system;

Step 5: if  $i < I-1$ , then set  $i := i+1$  and go to Step 2; otherwise stop.

Since it is possible to choose from various wavelet systems for synthesis and various densities for the scaling coefficients at the coarsest scale, we obtain a rich set of wavelet-based stochastic models.

## 2.2. Synthesis of Scaling Coefficients

In order to synthesize the scaling coefficients, we derive the second-order statistics of the scaling coefficients at any scale.

According to (27), we infer that

$$\begin{aligned} S_{i,k} &= \sum_{l_1} \sum_{l_2} h[k-2l_1] h[l_1-2l_2] S_{i-2,l_2} \\ &\quad + \sum_{l_1} \sum_{l_2} h[k-2l_1] g[l_1-2l_2] W_{i-2,l_2} \\ &\quad + \sum_{l_1} g[k-2l_1] W_{i-1,l_1} \end{aligned} \quad (12)$$

$$\begin{aligned} &= \sum_{l_1} \sum_{l_2} \sum_{l_3} h[k-2l_1] h[l_1-2l_2] \\ &\quad \times h[l_2-2l_3] S_{i-3,l_3} \\ &\quad + \sum_{l_1} \sum_{l_2} \sum_{l_3} h[k-2l_1] h[l_1-2l_2] \\ &\quad \times g[l_2-2l_3] W_{i-3,l_3} \\ &\quad + \sum_{l_1} \sum_{l_2} h[k-2l_1] g[l_1-2l_2] W_{i-2,l_2} \\ &\quad + \sum_{l_1} g[k-2l_1] W_{i-1,l_1} \end{aligned} \quad (13)$$

$$\begin{aligned} &= \sum_{n=2}^{\infty} \sum_{l_1} \sum_{l_2} \cdots \sum_{l_n} \left( \prod_{m=2}^{n-1} h[l_{m-1}-2l_m] \right) \\ &\quad \times h[k-2l_1] g[l_{n-1}-2l_n] W_{i-n,l_n} \\ &\quad + \sum_{l_1} g[k-2l_1] W_{i-1,l_1}. \end{aligned} \quad (14)$$

Thus, it follows that

$$\begin{aligned} E[S_{i,l} S_{i,k}] &= \sum_{n=2}^{\infty} \sigma_{i-n}^2 \sum_{l_1} \sum_{l_2} \cdots \sum_{l_n} h[l-2l_1] \\ &\quad \times h[k-2l_1] \left( \prod_{m=2}^{n-1} h[l_{m-1}-2l_m] \right)^2 \\ &\quad \times g^2[l_{n-1}-2l_n] \\ &\quad + \sigma_{i-1}^2 \sum_{l_1} g[l-2l_1] g[k-2l_1]. \end{aligned} \quad (15)$$

In our models, we choose to use biorthogonal  $B$ -spline wavelets whose synthesis filters are half-point

symmetric and hence satisfy

$$\sum_l h^2[2l] = \sum_l h^2[2l+1] = \frac{1}{2}, \quad (16)$$

$$\sum_l g^2[2l] = \sum_l g^2[2l+1] = \frac{1}{2}. \quad (17)$$

Therefore, it follows that

$$\begin{aligned} E[S_{i,l} S_{i,k}] &= \sum_{n=2}^{\infty} 2^{1-n} \sigma_{i-n}^2 \sum_{l_1} h[l-2l_1] h[k-2l_1] \\ &\quad + \sigma_{i-1}^2 \sum_{l_1} g[l-2l_1] g[k-2l_1] \end{aligned} \quad (18)$$

$$\begin{aligned} &= \frac{2^{4H-3}}{1-2^{2H-2}} 2^{-i(2H-1)} \sigma^2 \\ &\quad \times \sum_{l_1} h[l-2l_1] h[k-2l_1] \\ &\quad + 2^{2H-1} \cdot 2^{-i(2H-1)} \sigma^2 \\ &\quad \times \sum_{l_1} g[l-2l_1] g[k-2l_1]. \end{aligned} \quad (19)$$

Since

$$E[S_{i+2m,l} S_{i+2m,k}] = E[S_{i,l} S_{i,k}] \quad (20)$$

for any integer  $m$ , the process  $\{S_{i,k} : k \in \mathbf{Z}\}$  is wide-sense cyclostationary with period 2.

*Example.* If the Haar wavelet is used, i.e.,

$$h[n] = \frac{1}{\sqrt{2}}(\delta_n + \delta_{n-1}) \quad (21)$$

$$g[n] = \frac{1}{\sqrt{2}}(\delta_n - \delta_{n-1}), \quad (22)$$

then the process  $\{S_{i,k} : k \in \mathbf{Z}\}$  becomes wide-sense stationary due to that

$$E[S_{i,l} S_{i,k}] = \begin{cases} \frac{2^{2H-2}}{1-2^{2H-2}} 2^{-i(2H-1)} \sigma^2 & \text{if } l = k \\ \frac{2^{4H-3}-2^{2H-2}}{1-2^{2H-2}} 2^{-i(2H-1)} \sigma^2 & \text{if } |l-k| = 1 \\ 0 & \text{otherwise.} \end{cases} \quad (23)$$

### 2.3. Synthesis of Wavelet Coefficients

The synthesized process  $X_I(t)$  can be expressed as

$$X_I(t) = \sum_k S'_{I,k} \phi_{I,k}(t) \quad (24)$$

$$= \sum_k S'_{i_0,k} \phi_{i_0,k}(t) + \sum_{i=i_0}^{I-1} \sum_k W_{i,k} \psi_{i,k}(t) \quad (25)$$

where the biased scaling coefficients are given by

$$S'_{i+1,k} = S_{i+1,k} + 2^{-(i+1)/2} \mu \quad (26)$$

$$= \sum_l (h[k-2l] S'_{i,l} + g[k-2l] W_{i,l}) \quad (27)$$

To synthesize a non-negative process  $X_I(t)$ , we need to maintain the non-negativity of the biased scaling coefficients  $\{S'_{i,k} : i_0 \leq i < I, k \in \mathbf{Z}\}$ . To achieve this goal, we first choose  $S'_{i_0,k}$  to be non-negatively distributed. The probability density function (PDF) of  $S'_{i_0,k}$  can be log-normal, Rayleigh, Maxwell, gamma, etc. Secondly, we use a multiplicative model for synthesizing wavelet coefficients as

$$W_{i,k} = A_{i,k} S'_{i,k} \quad (28)$$

where  $\{A_{i,k} : k \in \mathbf{Z}\}$  are zero-mean, i.i.d. random variables for a fixed  $i$ , and  $A_{i,k}$  is independent of  $S'_{i,k}$  for any  $i$  and  $k$ .

We assume that the synthesis filters satisfy

- the support of  $g[n]$  is a subset of the support of  $h[n]$ ;
- $h[n]$  is non-negative, i.e.,  $h[n] \geq 0, \forall n$ .

The biorthogonal  $B$ -spline wavelets (including the Haar wavelet), whose two synthesis filters have the same even length, satisfy both conditions. Define

$$C_{h,g} = \min_{g[n] \neq 0} \frac{h[n]}{|g[n]|}. \quad (29)$$

In our models, the wavelet coefficients and the scaling coefficients satisfy

$$|A_{i,k}| \leq C_{h,g} \quad \forall i, k. \quad (30)$$

Using (27), we infer that

$$S'_{i+1,k} \geq \sum_l (h[k-2l] S'_{i,l} - |g[k-2l] W_{i,l}|) \quad (31)$$

$$\geq \sum_l (h[k-2l] S'_{i,l} - |g[k-2l]| C_{h,g} S'_{i,l}) \quad (32)$$

$$\geq 0. \quad (33)$$

The variance of the wavelet coefficient  $W_{i,k}$  is given by

$$E[W_{i,k}^2] = E[A_{i,k}^2] E[S_{i,k}^2]. \quad (34)$$

In our models, the PDF of  $A_{i,k}$  is chosen to be a mixture of two symmetric beta PDFs:

$$p_{A_{i,k}}(a) = \lambda_i p(a; q_{i,1}) + (1 - \lambda_i) p(a; q_{i,2}) \quad (35)$$

where  $0 \leq \lambda_i \leq 1$  and  $p(a; q)$  denotes the symmetric beta PDF with a shape factor  $q > 0$ , i.e.,

$$p(a; q) = \begin{cases} \frac{1}{C_q} (C_{h,g}^2 - a^2)^{q-1} & \text{if } |a| \leq C_{h,g} \\ 0 & \text{otherwise} \end{cases} \quad (36)$$

with the constant

$$C_q = 2^{2q-1} \int_0^{C_{h,g}} [x(C_{h,g} - x)]^{q-1} dx. \quad (37)$$



Since the variance associated with the PDF  $p(a; q)$  is  $C_{h,g}^2/(2q+1)$ , the variance of  $A_{i,k}$  is given by

$$E[A_{i,k}^2] = C_{h,g}^2 \left[ \frac{\lambda_i}{2q_{i,1} + 1} + \frac{(1 - \lambda_i)}{2q_{i,2} + 1} \right]. \quad (38)$$

### 3. FITTING THE MODELS TO DATA

For a given training data set, we determine the parameters of the proposed models from the empirical wavelet coefficients  $\{\widehat{W}_{i,k} : i_0 \leq i < I, k \in \mathbf{Z}\}$  and scaling coefficients  $\{\widehat{S}_{i_0,k} : k \in \mathbf{Z}\}$ .

#### 3.1. Least-Squares Estimation of $H$ and $\{\sigma_i^2\}$

We first compute  $\widehat{\sigma}_i^2$ , the variance of the empirical wavelet coefficients at scale  $2^i$  for  $i_0 \leq i < I$ . Then, we use the least-squares criterion to estimate  $H$  and  $\sigma^2$  from  $\{\widehat{\sigma}_i^2 : i_0 \leq i < I\}$  according to (6).

#### 3.2. Maximum Likelihood Estimation of the PDFs $\{p_{A_{i,k}}(a)\}$

In order to synthesize the wavelet coefficients at a fixed scale  $2^i$ , we need to determine the PDF  $\{p_{A_{i,k}}(a)\}$ . Due to (38), the mixing parameter  $\lambda_i$  can be expressed in terms of the desired variances of  $W_{i,k}$  and  $S'_{i,k}$ :

$$\lambda_i = \frac{(2q_{i,1} + 1)(2q_{i,2} + 1) \frac{\sigma_i^2}{E[(S'_{i,k})^2] C_{h,g}^2} - (2q_{i,1} + 1)}{2(q_{i,2} - q_{i,1})}. \quad (39)$$

We use a maximum likelihood criterion to estimate the shape parameters  $q_{i,1}$  and  $q_{i,2}$  from the empirical wavelet coefficients and scaling coefficients at scale  $2^i$ :

$$\max_{q_{i,1}, q_{i,2}} \prod_k p_{A_{i,k}} \left( \frac{\widehat{W}_{i,k}}{\widehat{S}'_{i,k}} \right). \quad (40)$$

### 4. SIMULATIONS

In our simulations, we use a measured traffic trace from the Bellcore ftp site [1]. We choose the Haar wavelet and model the biased scaling coefficients at scale  $2^{i_0}$  using the Rayleigh density.

Figure 1(a) and 1(b) depict a segment of the measured traffic data and a segment of the synthesized traffic data, respectively. Figure 1(c) and 1(d) illustrate the histograms of the measured trace and the synthesized trace, respectively. Figure 2(a) and 2(b) plot the autocovariance functions of the measured trace and the synthesized trace, respectively. Figure 2(c) and 2(d) plot the power spectra of the measured trace and the

synthesized trace, respectively. These figures demonstrate that the statistics of the measured traffic and the synthesized traffic are very close.

Figure 2(e) plots the probabilities of buffer overflow versus buffer size for a single-server queue fed with the measured trace and the synthesized trace. The figure shows that the queuing behaviors for the two traces are very similar.

### 5. CONCLUSION

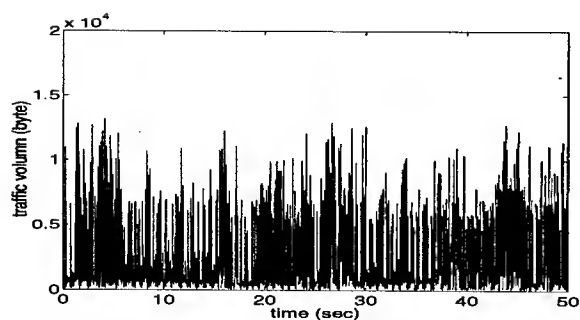
We have presented a set of wavelet-based stochastic models for  $1/f$  network traffic. Besides the accuracy shown in our simulations, our models possess the following features:

- parsimony: the model parameters include the parameters of the PDF of the scaling coefficients at scale  $2^{i_0}$ ,  $H$ ,  $\sigma^2$ , and  $\{q_{i,1}, q_{i,2} : i_0 \leq i < I\}$ ;
- computational efficiency: wavelet analysis and synthesis have low computational complexity.

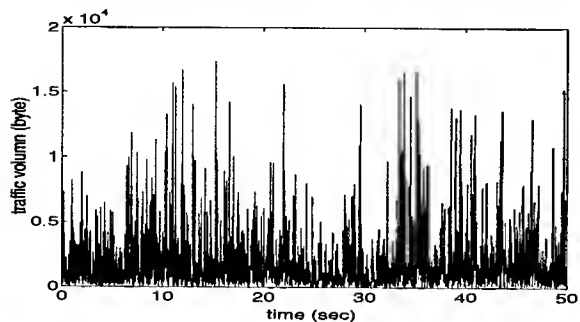
Therefore, the proposed models are very promising in network traffic engineering.

### REFERENCES

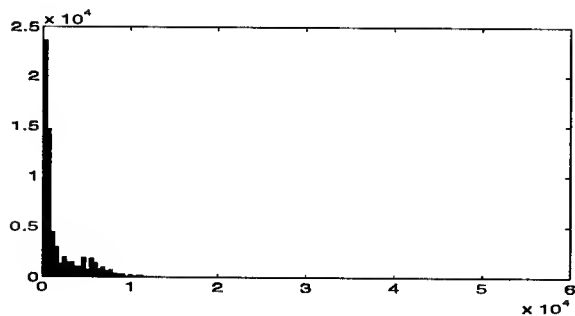
- [1] W. E. Leland, M. S. Taqqu, W. Willinger, and D. V. Wilson, "On the self-similar nature of Ethernet traffic (extended version)," *IEEE/ACM Trans. Networking*, vol. 2, pp. 1-15, Feb. 1994.
- [2] V. S. Frost and B. Melamed, "Traffic modeling for telecommunications networks," *IEEE Commun. Magazine*, vol. 32, pp. 70-81, Mar. 1994.
- [3] I. Daubechies, *Ten Lectures on Wavelets*. Philadelphia, PA: Soc. Indus. Appl. Math., 1992.
- [4] G. W. Wornell, *Signal Processing with Fractals: A Wavelet-Based Approach*. Upper Saddle River, NJ: Prentice-Hall, 1995.
- [5] S. Ma and C. Ji, "Modeling video traffic using wavelets," *IEEE Commun. Letters*, vol. 2, pp. 100-103, Apr. 1998.
- [6] R. H. Riedi, M. S. Crouse, V. J. Ribeiro, and R. G. Baraniuk, "A multifractal wavelet model with application to network traffic," *IEEE Trans. Inform. Theory*, vol. 45, pp. 992-1018, Apr. 1999.
- [7] S. G. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, pp. 674-693, July 1989.



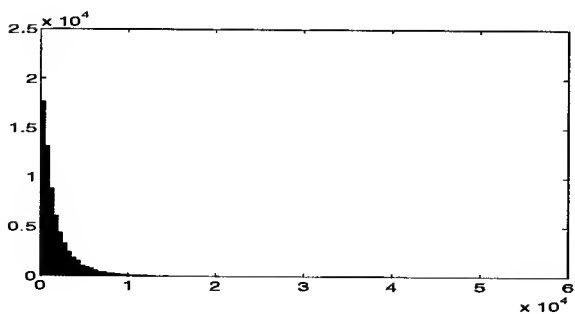
(a)



(b)

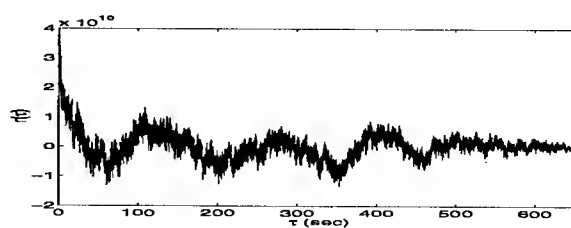


(c)

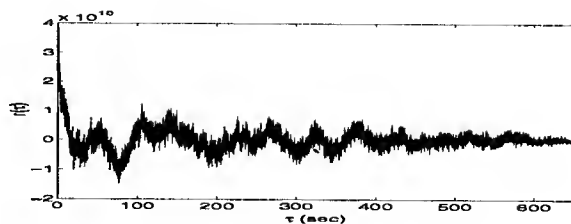


(d)

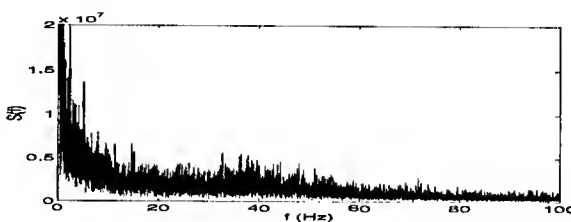
Figure 1: (a) A segment of the measured trace; (b) a segment of the synthesized trace; (c) histogram of the measured trace; (d) histogram of the synthesized trace.



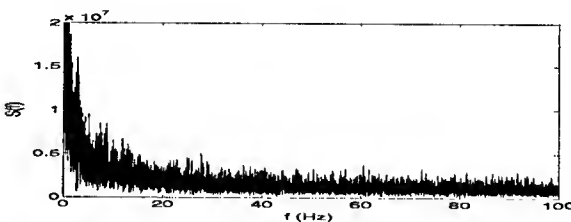
(a)



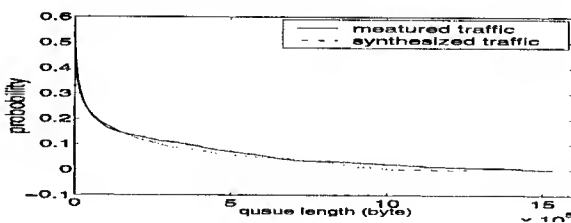
(b)



(c)



(d)



(e)

Figure 2: (a) Autocovariance function of the measured trace; (b) autocovariance function of the synthesized trace; (c) power spectrum of the measured trace; (d) power spectrum of the synthesized trace; (e) queuing behaviors of the two traces.

# THE EXTENDED ON/OFF PROCESS FOR MODELING TRAFFIC IN HIGH-SPEED COMMUNICATION NETWORKS

Xueshi Yang, Athina P. Petropulu and Vaughn Adams

Electrical and Computer Engineering Department,  
Drexel University, Philadelphia, PA 19104, USA  
Tel. (215) 895-2358 Fax. (215) 895-1695

## ABSTRACT

High-speed network traffic is impulsive and exhibits long-range dependence. While the latter characteristic has been studied extensively, the former has received much less attention. The On/Off model is a well known model for capturing the long-range dependence of traffic traffic. In this paper we propose an extension to the On/Off model, which allows the model to also capture the traffic impulsiveness. We provide queuing analysis of the proposed model, which along with numerical results suggests that the traffic marginal distribution may have significant impact on networking engineering.

## 1. INTRODUCTION

Extensive studies indicate that traffic in high-speed communication networks has self-similar [14], [4] and long-tailed characteristics [7], [6], [10]. There are many studies dealing with the self-similarity characteristic, the best known of which is the On/Off model [14]. In data communication networks, the packets are communicated in a "packet train" fashion; once a "packet train" is triggered, the probability that another packet will follow the current one is very large. The On/Off model is based on that packet train idea. A single source/destination active pair alternates between two states: the On, during which, there is data flow between source and destination, along either way, and the Off, which is the quiet duration. Both the On and Off durations follow a heavy-tail distribution. For heavy-tail phenomena the probability of large values decays hyperbolically instead of exponentially. The self-similar characteristics of the AFRP have been attributed to the heavy-tail properties of the On/Off states durations. However, in the seminal paper of [13] it was shown that the cumulative superposition of infinite AFRP's is fractional Brownian motion, which is Gaussian. This fact renders the superposition of AFRP's inconsistent with real traffic data, which is clearly non-Gaussian. In fact, the marginal distribution of a traffic flow can have a profound impact on network engineering, for example, it can significantly change queuing performance and buffer overflow probability [5]. In [5], it is shown that under different marginal distributions of the traffic streams, the packet loss rates differed by several orders of magnitude.

This work was supported by National Science Foundation under grant MIP-9553227

In this paper we propose the Extended On/Off process, as a way to overcome the limitation of the traditional On/Off model. Each user transmits or stays idle, with durations that are heavy-tail distributed, but, unlike the AFRP model, the bandwidth requirement during the transmission state is a heavy-tailed. We provide proofs for long-range dependence and heavy-tail properties of the proposed model for single user traffic and also for aggregated traffic. We provide analytical results on the queuing behavior of the proposed model, which indicate that the heavy-tailed reward process may affect queuing performance as much as the self-similar characteristics of the traffic flow. We also provide results based on real traffic to demonstrate the validity of our theoretical claims.

## 2. MATHEMATICAL PRELIMINARIES

A random variable is called *regularly varying* with index  $\alpha$ , to be denoted by  $X \in \mathcal{R}_\alpha$ , if for all  $k > 0$ ,

$$\lim_{t \rightarrow \infty} \bar{F}_X(kt)/\bar{F}_X(t) = k^\alpha. \quad (1)$$

A random variable with regularly varying distribution function is also referred to as *heavy-tail distributed*. The *Pareto* distribution is the simplest example of heavy-tailed distribution. Its survival function is given as:

$$\bar{F}_X(x) = \begin{cases} (\frac{k}{x})^\alpha, & x \geq k, \\ 1, & x < k, \end{cases} \quad (2)$$

where  $k$  is positive constant.

A random process  $x(t)$ , with finite second order statistics, is called *stationary process with long memory* [1], or *long-range dependence in the autocovariance sense*, if its autocovariance function decays hyperbolically as the lag  $k$  increases.

For processes which might not have second-order statistics, a structure measure different than the autocorrelation is needed. We will use the quantity defined in [12] i.e.,

$$I(\rho_1, \rho_2; \tau) \triangleq -\ln E\{e^{i(\rho_1 x(t+\tau) + \rho_2 x(t))}\} + \ln E\{e^{i\rho_1 x(t+\tau)}\} + \ln E\{e^{i\rho_2 x(t)}\} \quad (3)$$

and will be referring to the above quantity as the *generalized codifference*.

We will say that the stationary process  $X(t)$  is a *long-memory process in a generalized sense* if

$$\lim_{\tau \rightarrow \infty} -I(1, -1; \tau)/\tau^{\beta-1} = c \quad (4)$$

where  $c$  is some positive constant and  $\beta < 1$ . This definition of long-range dependence has been used for the first time in [11] to study the joint statistics of the power-law shot noise process.

## 2.1. The AFRP

The Alternating Fractal Renewal Process (AFRP), proposed in [14] for modeling of network traffic, is a process that alternates between two states, 0 or 1. The time  $\{X_n\}$ , spent in state 1, is a random variable with density function  $f_1(t)$ , and the time  $\{Y_n\}$ , spent in state 0, is a random variable with pdf of  $f_0(t)$ , where  $f_1(t)$ ,  $f_0(t)$  obey heavy-tailed distributions, i.e.

$$f_i(t) \sim t^{-(\alpha_i+1)}, \quad \text{where } i = 0, 1, \quad \alpha_i \in (1, 2) \quad (5)$$

Generally speaking,  $f_0(t) = f_1(t) = 0$  for  $t < 0$ , and the associated dwell mean times  $\mu_1 := E[X_n]$  and  $\mu_0 := E[Y_n]$  are finite. The expected value of the AFRP process  $X(t)$  is  $\mu_1/(\mu_0 + \mu_1)$ . The power spectral density of AFRP equals [9],

$$S(\omega) \triangleq E\{X(t)\}\delta(\omega/2\pi) + \frac{2\omega^{-2}}{\mu_0 + \mu_1} \operatorname{Re} \left\{ \frac{[1 - Q_0(-j\omega)][1 - Q_1(-j\omega)]}{1 - Q_0(-j\omega)Q_1(-j\omega)} \right\} \quad (6)$$

where  $Q_0(-j\omega)$ ,  $Q_1(-j\omega)$  are the Fourier transforms of  $f_0(t)$ , and  $f_1(t)$  respectively.

## 2.2. Related queuing results

We next summarize some queuing analysis results needed in the analysis of the proposed model.

Consider a GI/G/1 queue, and let  $w_n$  represent the actual waiting time of the  $n$ th arriving customer, which has a service time of  $\tau_n$ . Let  $W(\cdot)$  and  $B(\cdot)$  be the distribution functions of  $w_n$  and  $\tau_n$ , respectively. Also, let  $\sigma_{n+1}$  be the interarrival time between the  $n$ th and  $n+1$ th customer, and  $A(\cdot)$  be its distribution. In [3] it was shown that the distribution function of the stationary actual waiting time has a regularly varying tail if and only if the tail of the service time distribution varies regularly at infinity. In a form of a theorem, it was shown that [3]:

**Theorem 1** For GI/G/1 queuing system with traffic intensity  $\rho = b/a < 1$ , as  $t \rightarrow \infty$  it holds:

$$\bar{B}(t) = k(b/t)^{k+1}L(t) \iff \bar{W}(t) = \rho(1-\rho)^{-1}(b/t)^kL(t), \quad (7)$$

where  $\bar{B}(t) = 1 - B(t)$ ;  $\bar{W}(t) = 1 - W(t)$ ;  $a, b$  denote the mean of  $\tau_n$  and  $\sigma_n$  respectively;  $k > 0$  and  $L(t)$  is a slowly varying function.

In [8], Kella *et al* studied a storage model with a two-state random environment, that alternates between *down* and *up* states. During *down* times  $\{D_k : k \geq 1\}$ , there is net flow into the buffer according to a stochastic process,  $\{R_k(t) : t \geq 0\} : k \geq 1$ , and during *up* times  $\{U_k : k \geq 1\}$ , there is a flow out of the buffer at rate  $r$ . Let  $Q(t)$ ,  $t \geq 0$  denote the buffer content process at time  $t$ , and let

$$Q(\infty) \stackrel{d}{=} \lim_{t \rightarrow \infty} Q(t), \quad (8)$$

where  $\stackrel{d}{=}$  represents equality in distribution, and

$$Q_e \stackrel{d}{=} \lim_{n \rightarrow \infty} Q(t_n), \quad (9)$$

where  $\{t_n = D_1 + U_1 + \dots + D_n + U_n\}$ ,  $n \geq 1$ . The following theorem links the steady state buffer content distribution at continuous time  $t$ ,  $t \geq 0$  to that at time points  $t_n$ ,  $n \geq 1$  [8]:

**Theorem 2** For traffic intensity  $\rho := E[R_1(D_1)]/rE[U_1] < 1$ , the steady state buffer content distribution satisfies:

$$P[Q(\infty) > q] = \frac{u}{u+d} \rho P[Q_e + R_1(D_1)^* > q] + \frac{d}{u+d} P[Q_e + R_1(D_1^*) > q], \quad (10)$$

where  $u = E[U_1]$  and  $d = E[D_1]$ , with

$$P[R_1(D_1)^* > q] = \frac{1}{E[R_1(D_1)]} \int_q^\infty P[R_1(D_1) > s] ds, \quad (11)$$

and

$$P[R_1(D_1^*) > q] = \frac{1}{d} \int_0^{D_1} 1_{[R_1(t) > q]} dt, \quad (12)$$

where  $1_{[\cdot]}$  is the indication function. Both  $R_1(D_1)^*$  and  $R_1(D_1^*)$  are independent of  $Q_e$ .

## 3. THE EXTENDED AFRP (EAFRP)

In [13], it was shown that the aggregated cumulative version of many homogeneous or heterogeneous AFRP processes, is fractional Brownian motion, the only Gaussian process with stationary increments that is self-similar. However, the fact that the aggregated sum of AFRP is Gaussian is not consistent with the heavy-tail properties of high-speed network traffic, the tail index  $\alpha$  of which deviates far away from 2.

As a simple way to introduce impulsiveness in the overall traffic model, we here propose to treat the single-user bit rate as a random variable with heavy-tailed characteristics. Let us define the extended AFRP (EAFRP) process as follows:

(i) The On-periods  $\{X_n\}$ , and the Off-periods  $\{Y_n\}$  are i.i.d., independent of each other with distributions respectively  $F_1$  and  $F_0$ , and have finite mean  $\mu_1$  and  $\mu_0$ , respectively; (ii) The transmitting rates  $\{A_n\}$  during different on-periods are i.i.d. random variables with distribution function  $F_A$ , independent of  $\{X_n\}$  and  $\{Y_n\}$ , and have finite mean  $\mu_A$  (iii)  $F_1$ ,  $F_0$  and  $F_A$  are Pareto distributed, with tail indices respectively  $1 < \alpha_1$ ,  $\alpha_0$ ,  $\alpha_A < 2$ , and parameters  $k_1$ ,  $k_0$ ,  $k_A > 0$  respectively.

For a single AFRP it was shown in [9], that in the intermediate frequency range, the power spectrum follows a power-law function. We here examine the single AFRP in the frequency range around the origin, and show a result similar to that of [9], i.e.,

**Proposition 1** An AFRP with On and Off periods Pareto distributed with tail indices  $\alpha_1$  and  $\alpha_0$ , respectively, is long memory in the autocovariance sense.

The proof can be found in [16].

As the EAFRP is constructed based on the AFRP, it should also exhibit some long-range dependence. However, by letting the reward be heavy-tailed, the second order statistics are infinite. Thus, the long-range dependence of EAFRP will be studied in the generalized sense of (3).

**Proposition 2** Let  $E(t)$  be an EAFRP as defined above.

- For a fixed  $t$ ,  $E(t)$  is a heavy-tail random variable with tail index  $\alpha_A$ .
- $E(t)$  exhibits long-range dependence in the generalized sense. i.e.,

$$-I(1, -1; \tau) \sim c\tau^{1-\min\{\alpha_1, \alpha_0\}}, \quad c > 0 \quad (13)$$

Proof: see Appendix A.

The overall network traffic consists of the superposition of many single source/destination pairs. Thus, the proposed model for the overall traffic is the superposition of EAFRP's.

**Proposition 3** Let  $SE(t)$  be the superposition of  $M$  independent EAFRP's  $E_m(t)$ ,  $m = 1, \dots, M$ , with parameters denoted by a subscription e.g.  $\alpha_{1i}, k_{1i}, \alpha_{0i}, k_{0i}, \alpha_{Ai}, k_{Ai}$ .

- For some fixed  $t$ ,  $SE(t)$  is a heavy-tail random variable with tail index  $\min(\alpha_{A1}, \alpha_{A2}, \dots, \alpha_{AM})$ .
- $SE_M(t)$  has long-range dependence in the generalized sense

Proof: Can be found in [16].

#### 4. QUEUING ANALYSIS OF THE EAFRP MODEL

Let us Consider an EAFRP process feeding a stable queue. During the On state fluid enters in the queue, while during both the On and Off states fluid leaves the queue at constant rate  $r$ . For a stable queue,  $r$  is larger than the mean in-flow rate, i.e.

$$r > \frac{\mu_A \mu_1}{\mu_1 + \mu_0} \quad (14)$$

The buffer content, or queue length  $Q(t)$ , is a continuous-time stationary stochastic process. The case where  $A_n \equiv \text{constant} > r$ , and  $Y_n$  exponentially distributed has been extensively treated [2]. For the case of heavy-tailed  $A_n$  we propose the following result.

**Proposition 4** The steady state queue length of an EAFRP queue is heavy-tail distributed with tail index  $1 - \alpha_A \wedge \alpha_1$ , i.e.

$$P[Q(\infty) > q] \sim Cq^{1-(\alpha_A \wedge \alpha_1)}, \quad \text{as } q \rightarrow \infty, \quad (15)$$

where  $C$  is some constant independent of  $q$

*Proof:* In our queuing analysis we will employ the traditional methodology, where the distribution of the buffer content, or the queue length, is first computed at discrete time points and then the stationary distribution is derived.

Let us study the buffer content at time points:

$$\{t_n := \sum_{i=1}^n X_i + Y_i, \quad n = 1, 2, \dots\} \quad (16)$$

At those points the  $Q(t)$  satisfies the recursive equation:

$$Q_{n+1} = [Q_n + (A_n - r)X_n - rY_n]^+, \quad n = 1, 2, \dots, \quad (17)$$

where  $Q_n := Q(t_n)$ ,  $[\cdot]^+ := \max[0, \cdot]$ . We will assume that the fluid source begins with an on period, with empty buffer content at time zero, i.e.,  $Q_0 = 0$ .

The net input  $B_n$  during an on session is

$$B_n = (A_n - r)X_n, \quad n = 1, 2, \dots \quad (18)$$

Given that  $A_n$  and  $X_n$  are heavy-tail distributed, and for  $r < K_A$ ,  $B_n$  can be shown [16] to be heavy-tail distributed with tail index  $-\min(\alpha_A, \alpha_1)$ .

The queue length satisfies the same recursive equation as the successive waiting times in a GI/G/1 queue with service times  $\{(A_n - r)X_n\}$  and inter-arrival times  $\{rY_n\}$ ,  $n = 1, 2, \dots$ . Thus, applying (7) with  $\sigma = \alpha_A \wedge \alpha_1$ ,  $N(t) = C_1$ ,  $\beta = r\mu_0$ ,  $\alpha = (\mu_A - r)\mu_1$ , as  $q \rightarrow \infty$  we get:

$$P[Q_e > b] \sim \frac{C_1 r \mu_0}{r \mu_0 (\alpha_A \wedge \alpha_1 - 1) [(\mu_A - r)\mu_1 - r \mu_0]} b^{-(\alpha_A \wedge \alpha_1 - 1)} \quad (19)$$

Linking the On and Off periods to the down and up states of [8] we can apply Theorem 2 to get the the steady state queue length distribution as:

$$\begin{aligned} P[Q(\infty) > q] &= \frac{\mu_1}{\mu_1 + \mu_0} P[Q_e + R_1(X_1^*) > q] \\ &\quad + \frac{\mu_0}{\mu_1 + \mu_0} \rho P[Q_e + R_1(X_1)^* > q] \end{aligned}$$

with traffic intensity  $\rho = (\mu_A - r)\mu_1 / r\mu_0$ , and

$$P[R_1(X_1^*) > x] = \frac{1}{EX_1} E \int_0^{X_1} 1_{[(A_1 - r)t > x]} dt \quad (20)$$

and

$$P[R_1(X_1)^* > x] = \frac{1}{E(A_1 - r)X_1} \int_x^\infty P[(A_1 - r)X_1 > t] dt. \quad (21)$$

where  $R_1(X_1)^*$  and  $R_1(X_1^*)$  are independent of  $Q_e$ .

It can be shown that  $P[R_1(X_1)^* > x] \sim x^{1-\alpha_A \wedge \alpha_1}$ , and,  $P[R_1(X_1^*) > x] \sim x^{1-\alpha_1}$ , as  $x \rightarrow \infty$ . Combining (19) yields that the stationary queue length distribution is heavy-tail distributed with tail index  $1 - \alpha_1 \wedge \alpha_A$ .  $\square$

So, driven by a single EAFRP source, when the marginal distribution of the transmitting rates has a heavier tail than the on periods, the buffer content distribution will be significantly changed, namely the asymptotic tail index. Furthermore, from actual high-speed LAN traffic measurement, it is observed that in most cases, the transmitting rates' tails are much heavier than the on-periods'. It implies that, in such cases, the asymptotic queue length behavior is determined by the marginal distribution, instead of that of on periods.

## 5. EXPERIMENTS

In this section, we first validate our claim that in actual single user traffic the transmitting rates are heavy-tailed distributed. We then demonstrate that the EAFRP indeed can model traffic in high-speed communication networks. Finally a numerical queuing simulation is performed to validate our theoretical finding according to which the traffic marginal distribution can affect the queuing performance as much as the traffic long-range dependence.

Our real traffic data was obtained from the 100-Mbps high-speed Ethernet network at the Electrical and Computer Engineering Department, Drexel University. It contains all packets transmitted to and from a Unix server in 3 continuous days. Out of the total traffic we separated the flow between a single user (han.ece.drexel.edu) and the server (cbis.ece.drexel.edu), which occurred on May 19th 2000 from 8:20AM to 21:00PM. The EAFRP was formed by firstly choosing an appropriate threshold value. If no packets were transmitted during a time period longer than the threshold value, that period was considered to be an Off state. The Off state was followed by an On state, which started as soon as packet activity resumed. The transmitting rate during each On period was calculated by averaging the total bytes transmitted during that period over the period duration.

To determine the presence or absence of the heavy-tail effects, the most commonly used methods are the log-log complementary distribution (LLCD) graph and the Hill estimator [14]. For  $t_{th} = 0.1$  sec., Fig. 1 depicts the Hill estimate plot of the transmitting rates during different on periods. The heavy-tailness of the transmitting rates is revealed by the stable Hill estimator plot. The same plot also shows the tail index to be close to 1.

Next, we proceed to use EAFRP model to synthesize the traffic. The tail indices of the EAFRP,  $\alpha_1$ ,  $\alpha_0$  and  $\alpha_A$  were obtained through the Hill estimator applied on the real data. The cutoffs, i.e.  $k_1$ ,  $k_0$  and  $k_A$  were set to be the minimum values of the On/Off durations and transmitting rates, respectively. These parameters were found to be:  $\alpha_1 = 1.5$ ,  $k_1 = 1$ ,  $\alpha_0 = 1.2$ ,  $k_0 = 10$ ,  $\alpha_A = 1.0$ ,  $k_A = 30$ . Figures 2(a) and (b) illustrate the actual traffic and the synthesized EAFRP. We observe that the outlook of the two traces are very alike, i.e. they are very impulsive. To affirm this "visual check", we plot the LLCD of the both traces in (c) and (d) respectively. The linearity in both plots indicate that both traces are indeed heavy-tail distributed. A further check of the similarities of these two traces is done by estimating their generalized codifferences, which are shown in (e) and (f) respectively. It is obvious that both data traces exhibit the same kind of long-range dependence in the generalized sense.

In the following, we performed a simple numerical simulation of a stable queue fed by an EAFRP process, of which the tail index of the On state and transmitting rate were 1.5, and 1.3 respectively. Other parameters were taken to be  $\alpha_0 = 1.3$ ,  $k_1 = k_0 = 1$ ,  $k_A = 10$ . The server service rate was set to 46 corresponding to a traffic intensity 38%. Based on 20 Monte Carlo simulations of time length  $10^6$  seconds we estimated the complementary mean queue length, which is shown in Fig. 3. The slope of a line, which was fitted to the queue length in the least-squares sense, was found to be 0.2603, which is very close to the theoretical value of

0.3. As expected by our theoretical results, the marginal distribution becomes the dominant factor in determining the queuing performance, which can have a profound effect in self-similar traffic engineering.

## 6. CONCLUSIONS

In this paper, we proposed the EAFRP model for modeling single user traffic in high-speed data networks. Both theoretical and simulations indicate that the EAFRP model is able to capture the impulsiveness as well as the long-range dependence of traffic. Our model can be easily configured. It has only 6 parameters, which can be used to produce versatile desired traffic flow traces. In a scaled network environment the total traffic at any load can be synthesized as the superposition of EAFRPs, where the number of EAFRPs corresponds to the active source/destination pairs in the whole network. The EAFRP model revealed an intriguing result in traffic engineering. Contrary to what has been assumed so far, the Hurst parameter is not the only factor in determining buffer dimensioning and loss-rate estimation. Both our analytical queuing results and experiments indicated that the traffic marginal distribution is equally important to the self-similarity, which in turn suggests that the marginal distribution should be taken into account in network infrastructure design.

## REFERENCES

- [1] J. Beran, *Statistics for Long-Memory Processes*, Chapman & Hall, New York, 1994.
- [2] O.J. Boxma and V. Dumas, "Fluid queues with long-tailed activity periods distributions", special Issue on *Stochastic Analysis and Optimization of Communication Systems* of the journal *Computer Communications*, 1998.
- [3] J.W. Cohen, "Some results on regular variation for distributions in queuing and fluctuation theory", *J. Appl. Probab.*, Vol.10:343-353, 1973.
- [4] M. E. Crovella, A. Bestavros, "Self-similarity in world wide web traffic: Evidence and possible causes", *IEEE/ACM Trans. Networking*, Vol.5, No.6, December 1997.
- [5] M. Grossglauser and J. Bolot, "On the relevance of long-range dependence in network traffic", *IEEE/ACM Trans. Networking*, Vol.7, No.5, October 1999.
- [6] J. Ilow, "Forecasting Network Traffic Using FARIMA Models with Heavy Tailed Innovations", *ICASSP #000*, Istanbul, Turkey, June 2000.
- [7] A. Karasavardis, D. Hatzinakos, "A non-Gaussian self-similar process for broadband heavy traffic modeling", *GOLBECOM'98*, Sydney, Australia.
- [8] O. Kella, W. Whitt, "A storage model with a two-state random environment", *Operations Research*, Vol.40, Supp. No.2:S257-S262, May-June, 1992.
- [9] S. B. Lowen and M. C. Telch, "Fractal renewal processes generate 1/f noise", *Physical Review E*, Vol.47, No.2, Feb 1993.
- [10] T. Mikosch, S. Resnick, H. Rootzen and A.W. Stegeman, "Is Network Traffic Approximated by Stable Levy Motion or Fractional Brownian Motion?", technical report, Cornell University, 1999.
- [11] A.P. Petropulu, J.-C. Pasquet, X. Yang, "Power-law shot noise and relationship to long-memory processes", *IEEE Trans. on Sig. Proc.*, July 2000.
- [12] G. Samorodnitsky and M. S. Taqqu, *Stable Non-Gaussian Random processes: Stochastic Models with Infinite Variance*, New York: Chapman and Hall, 1994.
- [13] M.S. Taqqu, W. Willinger and R. Sherman, "Proof of a fundamental result in self-similar traffic modeling", *Computer Communication Review* #7, pp.5-23, 1997.
- [14] W. Willinger, M.S. Taqqu, R. Sherman, and D.V. Wilson, "Self-similarity through high-variability: statistical analysis of Ethernet LAN traffic at the source level", *IEEE/ACM Trans. Networking*, Vol.5, No.1, February 1997.
- [15] X. Yang, A.P. Petropulu, and V. Adams, "The extended on/off model for high-speed data networks", *Applications of Heavy Tailed Distributions in Economics, Engineering and Statistics*, Juen 3-5, Washington DC, 1999.
- [16] X. Yang, A.P. Petropulu, and V. Adams, "The extended on/off process for modeling traffic in high-speed communication networks", *IEEE Trans. on Sig. Proc.*, submitted in 2000.

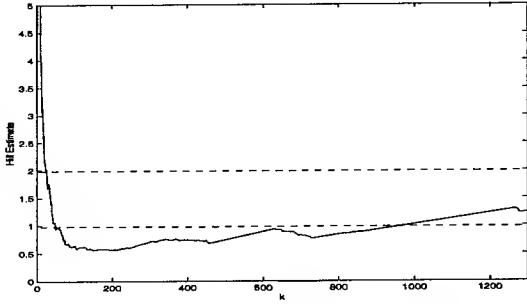


Figure 1: The Hill estimator the transmitting rates of actual single user traffic.

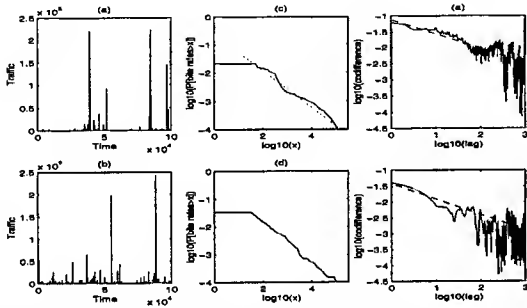


Figure 2: The actual single user network traffic (1st row), the synthesized traffic (2nd row) and their corresponding LLCD plots and codifference estimate.

## A. APPENDIX A

Let  $E(t) = A(t)V(t)$ , where  $V(t)$  is an AFRP and  $A(t)$  represents the random transmission rate. The density function of  $E(t)$  equals:

$$f_E(e) = P[V(t) = 0]\delta(e) + P[V(t) = 1]f_A(e) \quad (22)$$

where  $\delta(e)$  is the Dirac function, taking value of 1 at 0. Thus,  $f_E(e)$  is a scaled version of  $f_A(e)$ , which is a power-law function. Hence,  $E(t)$  is heavy-tail random variable with tail index  $\alpha_A$  for fixed  $t$ .

To show generalized long-range dependence we proceed as follows. It holds that:

$$E\{e^{s_1 E(t+\tau) + s_2 E(t)}\} = E\{\Phi_V(s_1 A(t+\tau), s_2 A(t))\} \quad (23)$$

For notational convenience, let  $V_1 = V(t+\tau)$ ,  $V_2 = V(t)$ ,  $A_1 = A(t+\tau)$ ,  $A_2 = A(t)$

$$\begin{aligned} \Phi_V(s_1, s_2) &= E\{e^{s_1 V_1 + s_2 V_2}\} \\ &= 1 + s_1 E\{V_1\} + s_2 E\{V_2\} + \frac{1}{2}(s_1^2 E\{V_1^2\} \\ &\quad + s_2^2 E\{V_2^2\} + 2s_1 s_2 E\{V_1 V_2\}) + \dots \end{aligned} \quad (24)$$

Taking into account that  $E\{V_1^n V_2^m\} = E\{V_1 V_2\}$  and also the stationarity of  $V(t)$  we find that:

$$\Phi_x(s_1, s_2) = 1 + (e^{s_1} + e^{s_2} - 2)\eta + (e^{s_1} - 1)(e^{s_2} - 1)E\{V_1 V_2\} \quad (25)$$

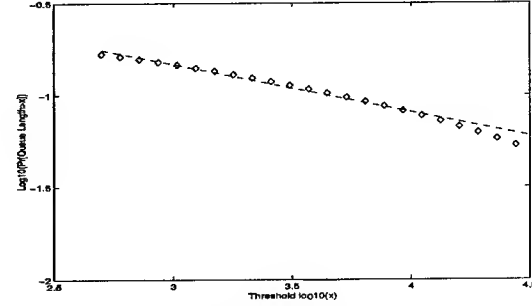


Figure 3: Complementary queue length distribution for EAFRP input, with Pareto distributed transmitting rates of tail index 1.3. On and Off periods are Pareto distributed of tail indices 1.5 and 1.3, respectively.

where  $\eta = E\{V(t)\}$ .

Now plugging (25) and (23) in the generalized codifference expression (see eq. (3)) will yield terms like  $\Phi_A(s_1)$ ,  $\Phi_A(s_2)$ ,  $\eta$  and also the term  $E\{e^{s_1 A_1 + s_2 A_2}\}$ . The latter term depends on whether  $t + \tau$  and  $t$  are in the same or different states. Let us denote the residue life of the state at time  $t$  by  $T$ . Then,

$$\begin{aligned} E\{e^{s_1 A_1 + s_2 A_2}\} &= E\{e^{s_1 A_1 + s_2 A_2} | T < \tau\} P\{T < \tau\} \\ &\quad + E\{e^{s_1 A_1 + s_2 A_2} | T > \tau\} P\{T > \tau\} \\ &= \Phi_A(s_1)\Phi_A(s_2)P\{T < \tau\} + \Phi_A(s_1 + s_2)P\{T > \tau\} \end{aligned}$$

From basic renewal theory, we have

$$\begin{aligned} P\{T > \tau\} &= P\{T > \tau | t \in \text{On state}\}P\{t \in \text{On state}\} \\ &\quad + P\{T > \tau | t \in \text{Off state}\}P\{t \in \text{Off state}\} \\ &= \frac{k_1^{\alpha_1} \tau^{1-\alpha_1}}{(\mu_1 + \mu_0)(\alpha_1 - 1)} + \frac{k_0^{\alpha_0} \tau^{1-\alpha_0}}{(\mu_1 + \mu_0)(\alpha_0 - 1)} \end{aligned}$$

Also, from proposition 1, we have

$$E\{V_1 V_2\} \stackrel{\tau \rightarrow \infty}{\sim} \eta^2 + c\tau^{1-\alpha_i} \quad (26)$$

Considering the approximation  $\log(1+x) \sim x$ ,  $|x| < 1$  and for  $\tau \rightarrow \infty$ :

$$\begin{aligned} I(s_1, s_2; \tau) &= -\ln \left[ ((\Phi_A(s_1) - 1)\eta + 1)((\Phi_A(s_2) - 1)\eta + 1) \right. \\ &\quad \left. + (\Phi_A(s_1) - 1)(\Phi_A(s_2) - 1)c\tau^{1-\alpha_i} + O(\tau^{2(1-\alpha_i)}) \right] \\ &\quad + \ln[1 + \eta(\Phi_A(s_1) - 1)] + \ln[1 + \eta(\Phi_A(s_2) - 1)] \\ &= -\frac{(\Phi_A(s_1) - 1)(\Phi_A(s_2) - 1)c}{((\Phi_A(s_1) - 1)\eta + 1)((\Phi_A(s_2) - 1)\eta + 1)} \tau^{1-\alpha_i} \\ &\quad + O(\tau^{2(1-\alpha_i)}) \\ &\sim c_2 \tau^{1-\alpha_i} \end{aligned} \quad (27)$$

where  $c_2$  is some constant and  $\alpha_i = \min\{\alpha_1, \alpha_0\}$ . Setting  $s_1 = j$ ,  $s_2 = -j$  it is easy to find that the discriminant of the denominator equals  $-4E\{\sin^2(A)\} < 0$ . Thus, the denominator is always positive, and as a result,  $c_2$  is always negative, which completes the proof.

# A SIMULATION STUDY OF THE IMPACT OF SWITCHING SYSTEMS ON SELF-SIMILAR PROPERTIES OF TRAFFIC

*Yunkai Zhou and Harish Sethu*

Department of ECE, Drexel University  
3141 Chestnut Street  
Philadelphia, PA 19104-2875.  
{kenty, sethu}@ece.drexel.edu

## ABSTRACT

Recent research has shown that traffic in Ethernet and other networks tends to exhibit properties of self-similarity such as long-range dependence and a high degree of correlation between arrivals. This paper investigates the impact of the switching network on the self-similar properties of the traffic. This simulation study reveals that switching networks tend to reduce the self-similarity of highly self-similar traffic. This is because of the truncation of long bursts due to packet discards, and also because of aggregation of flows through concatenated rather than superposed bursts. On the other hand, switching systems have the opposite effect of increasing the self-similarity of input traffic that has no self-similar properties such as traffic with Poisson or uniformly random distributions. This paper also presents simulation-based evidence of the causes behind these phenomena.

## 1. INTRODUCTION

Recent work by many researchers has shown that traffic in Ethernet and other networks tends to be bursty at many or all time-scales [2, 6], and that this phenomenon can be mathematically described using the notion of self-similarity. Extensive research has been done on the impact of the self-similar properties of traffic on network design issues, such as queueing performance [10], switch performance [3], congestion control [5] and scheduling algorithms [4]. While it is clear that traffic characteristics have an impact on network design issues, it is also true that the properties of a network have an impact on the characteristics of the traffic as it progresses through the network. Very few studies, however, have addressed this issue of changes in traffic characteristics caused by the network [1, 8, 12, 14]. These studies have focused on only the impact of individual components of a network such as the traffic shaper [12], the packet scheduler [1, 14] or a single-server queue [8], as opposed

to the impact of the entire network as a whole. In addition, studies such as [8] have obtained insightful theoretical results which, however, cannot be readily applied to realistic network environments to solve problems in network engineering. Further, studies such as in [12, 14] only consider short-range burstiness, which does not capture all of the features of self-similar traffic, especially long-range burstiness as observed in [2, 6].

This paper presents a simulation study of the impact of a switching network on the self-similar properties of the traffic, and investigates the causes underlying the observed phenomena. We use self-similar traffic generated using the fractional ARIMA model [7], and a baseline Banyan topology for the switching network. Section 2 discusses the network and the traffic model in greater detail.

Our simulation study reveals that switching networks tend to reduce the self-similarity of highly self-similar traffic. This is because of the truncation of long bursts due to packet discards, and also because of the aggregation of flows through concatenated rather than superposed bursts. On the other hand, switching systems also increase the self-similarity of input traffic that has no self-similar properties such as traffic with Poisson or uniformly random distributions. Section 3 presents these simulation results and the related analysis with simulation-based evidence of the causes behind the phenomena that yield these results. This section also explains our results in relation to those obtained in [8] and [11]. Section 4 concludes the paper.

## 2. NETWORK AND TRAFFIC MODEL

### 2.1. Network Model

This study uses an  $N \times N$  baseline Banyan multistage network, with  $N$  source nodes and  $N$  destination nodes. The switching network consists of  $\log_m N$  stages of  $m \times m$  switching elements. In Banyan topologies, the path between a source end-point and a destination end-point is unique. This property of Banyan networks helps our study of the

This work was supported in part by U.S. Air Force Contract F30602-00-2-0501



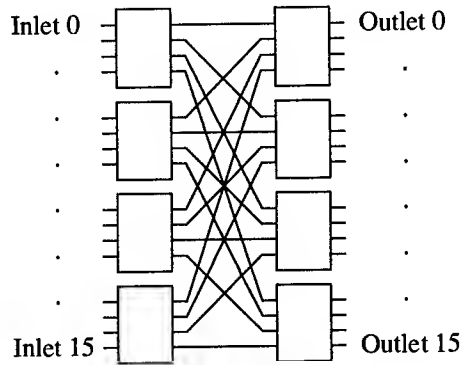


Figure 1: Banyan network with  $N = 16$  and  $m = 4$ .

impact of switching systems, since it eliminates other secondary effects such as due to the choice of a routing algorithm. The popularity of Banyan topologies in real implementations is an additional motivation behind our use of this network model. Figure 1 shows a baseline Banyan network topology with  $N = 16$  and  $m = 4$ .

Each source node consists of  $N$  traffic generators, each of which generates traffic intended for a distinct destination node. Thus, the system consists of a total of  $N^2$  traffic generators. In our simulations, traffic generators are all independent, and generate no more than one packet per cycle. We assume that packet lengths are constant, and that exactly one packet can be transmitted during each cycle across any port. If more than one packet are created in a source node during the same cycle, only one of these is allowed to be transmitted while all the others are buffered in a queue. We assume that the queue sizes at the source nodes are large enough that no packet is ever dropped before it enters the network. Destination nodes drain packets from the output ports of the last stage of switching elements, at the maximum rate of one packet per cycle.

In the switching elements, each input port is associated with an input buffer of a fixed small capacity of 4 packet lengths. Each output port contains a dedicated output buffer. In addition, our simulations also use a shared output buffer of capacity equivalent to 4 packets per output port for additional space for the output queues. Under most traffic conditions, the shared buffer improves performance through better buffer utilization. During each cycle in our simulations, switching elements can accept no more than one packet at each input port into the input queue. Each non-empty output queue transmits exactly one packet to the output port in each cycle. We use the round-robin scheduling algorithm to transfer packets to and from the shared queue. A packet arriving at an input port first enters the associated input buffer, then the shared output buffer, and finally the output buffer corresponding to the destination port. Our model ensures that the maximum bandwidth with which the shared buffer

can be written into or read from, is equal to the maximum aggregate input or output bandwidth of the switch. Packets arriving at a full input buffer are dropped. No packets, however, are dropped at any other point within the switching element, i.e., packets are forwarded to the shared buffer, or to an output buffer only if there is room available.

## 2.2. Traffic Model

We use the fractional autoregressive integrated moving average (FARIMA) model [7] to synthesize self-similar traffic. FARIMA( $p, d, q$ ) is defined as

$$\Phi(B)X_n = \Theta(B)\Delta^{-d}\epsilon_n,$$

where  $B$  is the backward operator, i.e.,  $Bx_n = x_{n-1}$ . The definition above can be also expressed as

$$X_i = \Delta^{-d}\epsilon_i - \theta_1\Delta^{-d}\epsilon_{i-1} - \dots - \theta_q\Delta^{-d}\epsilon_{i-q} + \phi_1X_{i-1} + \dots + \phi_pX_{i-p}. \quad (1)$$

In equation (1),  $\Delta^{-d}$  is defined as  $\Delta^{-d} = \sum_{i=0}^{\infty} b_i(-d)B^i$ , where  $b_0(-d) = 1$  and

$$b_i(-d) = \frac{\Gamma(i+d)}{\Gamma(d)\Gamma(i+1)}, i = 1, 2, \dots$$

When the innovation  $\epsilon_i$  is a stable process with index  $\alpha$ , i.e.,  $\epsilon_i \sim S_\alpha(\sigma, \beta, \mu)$ , the Hurst parameter,  $H$ , and the quantities  $\alpha$  and  $d$  are related by  $d = H - 1/\alpha$ . In this paper, we use the Hurst parameter as the measure of the degree of self-similarity. The Hurst parameter has a range of  $0.5 \leq H \leq 1$ , and a larger value of  $H$  implies a higher degree of self-similarity. Throughout our work, we use  $\alpha = 1.2, \sigma = 1, \beta = 0, \mu = 0, p = 50$ , and  $q = 400$ .  $\Theta(B)$  is generated by selecting  $\theta_i$  in  $[0, 0.05]$  randomly and independently. Unlike  $\Theta(B)$ ,  $\Phi(B)$  is generated by selecting  $p/2$  complex roots and their conjugates, since  $X_i$  converges only if all roots of  $\Phi(B)$  are in the unit circle. The real and imaginary components of each root are uniform in  $[0, 0.05]$ . Finally, we normalize  $X_i$  to a series of 1's or 0's indicating whether or not a packet is generated during a given cycle.

The variance-time plot [9] is used to estimate the Hurst parameter of observed network traffic. For a self-similar time series  $X(k)$ ,  $X_m(k)$  is defined as

$$X_m(k) = \frac{1}{m} \sum_{i=mk}^{(m+1)k-1} X(i),$$

and

$$\text{var}X_m = \text{var}X/m^\beta,$$

where  $H = 1 - \beta/2$ . Taking the logarithm of the equation above, we get,

$$\log \text{var}X_m = -\beta \log m + \log \text{var}X,$$

From the above equation, the Hurst parameter is determined by the slope of the plot of  $\log \text{var}X_m$  vs.  $\log m$ .

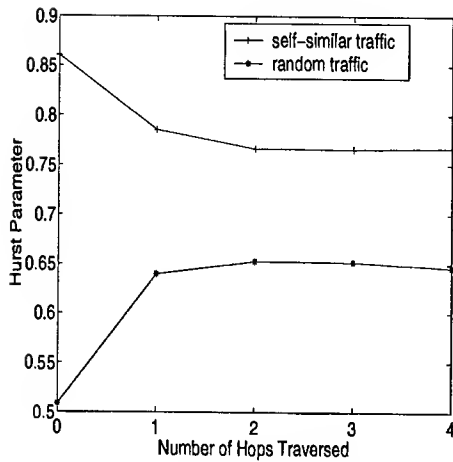


Figure 2: Per-hop changes in self-similarity.

### 3. SIMULATION RESULTS AND ANALYSIS

In a switching element with buffers, a flow typically consumes more space during a bursty period. Under such conditions, depending on the buffer sharing policy and the buffer sizes, either other flows suffer from less empty space, or the bursty flow suffers a higher packet loss rate. In either of these cases, the traffic characteristics change due to delays or losses or both. If two or more flows are bursty at the same time, these effects are further magnified. This section presents our study of these effects on the self-similar properties of traffic.

Our study includes two kinds of traffic sources, self-similar and uniform random traffic. A uniform random traffic source generates a packet during each cycle with a certain probability  $p$ , with uniformly distributed packet destinations. Like Poisson traffic, uniform random traffic has a Hurst parameter of 0.5, indicating that it has no self-similar properties.

Our simulation study shows that the impact of a switching system on the self-similarity of the traffic depends on the self-similarity of the input traffic itself. A switching system reduces the self-similarity of highly self-similar traffic, while it increases that of non-self-similar traffic such as uniform random traffic. Figure 2 illustrates this phenomenon of the opposite nature of the effects observed depending on the self-similarity of the input traffic itself. When the traffic is uniformly random, the Hurst parameter increases from 0.5 to 0.64 after the first stage and stays around 0.65 thereafter. When the input traffic has a high level of self-similarity, the Hurst parameter drops from 0.86 to 0.78 after the first stage, and further to 0.76 after the second stage. This interesting phenomenon shows us that, switching networks have the effect of shaping the traffic characteristics to a moderate level of self-similarity. In the following, we investigate

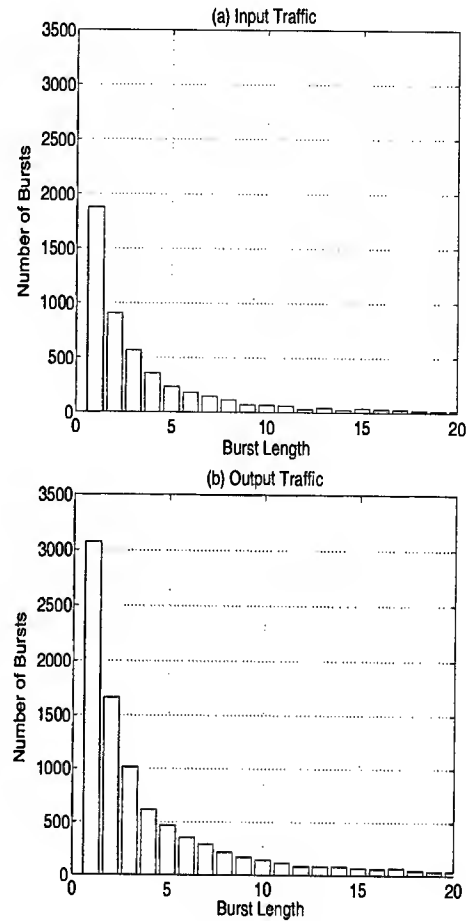


Figure 3: Distribution of short burst length, (a) input traffic and (b) output traffic.

and present simulation-based evidence of the causes of this phenomenon.

In the case of uniform random traffic, the probability of packet arrivals during each cycle is independent of the packet arrival pattern during the previous cycles. As the traffic progresses through a switching network with buffers in the switching elements, this independence assumption progressively becomes less valid. Packets for the same output port that independently arrive at different times, due to congestion, end up waiting in the buffers for transmission, and get transmitted in a burst at the output port. This phenomenon adds burstiness to the traffic at each new hop in the path of the traffic, changing the output traffic characteristics to something other than random uniform traffic. Packet arrivals at subsequent hops of the network are now correlated, as reflected in the increased Hurst parameter of the traffic.

The distribution of burst lengths in highly self-similar traffic is heavy-tailed, i.e., the probability distribution is given by  $P[X > x] \sim x^{-\alpha}$ . Such a distribution decays

Number of Nodes	$H$ (Input Traffic)	$H$ (Output Traffic)	Percentage Decrease
$2 \times 2$	0.860	0.864	$\sim 0$
$4 \times 4$	0.859	0.843	2%
$8 \times 8$	0.863	0.812	6%
$16 \times 16$	0.861	0.763	11%

Table 1: Self-similarity of traffic vs. number of nodes.

more slowly than exponentially, causing a high likelihood of long bursts in self-similar traffic. However, because of congestion and limited buffering capacities in the switching elements, long bursts do not easily survive in switching networks. In fact, long bursts self-destruct through causing congestion, triggering discarding of packets and thus breaking the long burst into smaller ones. For example, in our simulations, the longest burst observed at the traffic source had more than 6,000 consecutive packets, while the output traffic had no bursts longer than 900 packets. This is also illustrated in Figure 3, which shows that the distribution of short burst lengths of the input and the output traffic of the network. The output traffic has a larger percentage of shorter bursts, and with a significantly smaller average burst length. Note that the relative increase in shorter bursts increases the value of the index  $\alpha$  in a heavy-tailed distribution given by  $P[X > x] \sim x^{-\alpha}$ . This, in turn, has the effect of reducing the Hurst parameter since, as shown in [13],  $H = (3 - \alpha)/2$ .

In addition to the reduction in burst lengths, the other reason for this phenomenon is that aggregation of flows in networks typically has the effect of reducing variation over larger scales. It should be understood that this phenomenon is quite different from that shown in [11]. Willinger *et al.* show that a superposition of many *ON/OFF* traffic sources exhibits properties of self-similarity when the lengths of the *ON* and *OFF* periods are independent and follow a heavy-tailed distribution. Because of the limited bandwidth of output links, a true superposition is never possible in switching networks. Bursts are actually concatenated rather than superposed on top of each other on the output links. A superposition increases variation across scales, but a concatenation actually has the effect of spreading out the peaks and thus smoothening out the variations.

The phenomenon discussed above can be verified through simulation using a single  $m \times m$  switching element and varying  $m$ . In this model, each output link is fed by  $m$  self-similar traffic sources at the inputs. Table 1 shows the impact on the self-similarity of the output traffic for different values of  $m$ . As can be observed from Table 1, the self-similarity of traffic decreases as the level of aggregation increases. This reduction in the self-similarity of traffic with aggregation, is also the reason that highly self-similar traffic

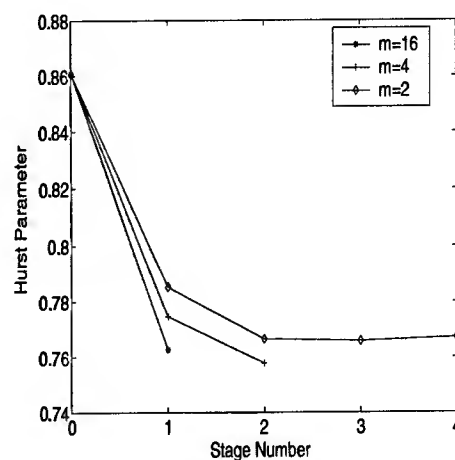


Figure 4: Per-hop changes for different switch sizes.

reduces in self-similarity as it progresses through each hop in the switching network. This is easily understood from noting that the output links further hops away from the traffic sources carry more of an aggregated traffic than the ones closer to the sources.

The same phenomenon is apparent in the impact of the size of switching elements used in the topology of a switching network. A network designed using  $2 \times 2$  switching elements, as compared to  $4 \times 4$  switching elements, will contribute to a smaller decrease in the observed Hurst parameter after the first hop. A network with  $4 \times 4$  switching elements achieves the same level of aggregation in fewer hops than one using  $2 \times 2$  switching elements. Figure 4 shows the per-hop changes in the self-similarity of traffic for switching networks with different sizes of switching elements.

It is worthwhile to discuss our results in relation to those obtained by Song *et al.* [8] in their study of self-similarity of output traffic at a single server with an infinite buffer. It is proved in [8] that if the queue length has finite variance, the self-similar properties of input and output traffic remain the same. In fact, it is shown that both the input and output traffic have the same Hurst parameter. Noting that it is unrealistic to assume an infinite buffer, the authors in [8] argue that in real switches, the condition that queue length has finite variance is always satisfied. However, another important impact of finite buffers should be considered—traffic can be accepted into the buffer only if there is available space. In the absence of a feedback mechanism, packet discarding becomes inevitable which changes the traffic characteristics; in the presence of a feedback mechanism such as credit-based flow control, the characteristics of arriving traffic itself changes. A second important reason for the apparent discrepancy between our results and that in [8] is that our results use the self-similar properties of *aggregate* traffic at each output link, while the results in [8] compare

the properties of *individual* flows before and after service by the server. Our approach to only analyze aggregate traffic is motivated by the fact that most switches and routers do not maintain per-flow queueing, and therefore, only the characteristics of the aggregate traffic at each input or output link is important to the performance and related issues in the design of switches and routers.

All of the results presented in this paper were obtained at moderate or heavy traffic loads. As one might expect, the impact of the switching network on the traffic characteristics is minimal when the traffic load is small.

#### 4. CONCLUSION

In this paper, we have presented simulation studies that show that highly self-similar input traffic reduces in its self-similarity as it progresses through a switching network. Our analysis indicates that this phenomenon is caused by the truncation of long bursts due to packet discarding, and by the aggregation of flows through concatenation of bursts. On the other hand, during periods of congestion in networks with buffers, input traffic with no self-similar properties increases in self-similarity as it progresses through the network.

These results have important implications relevant to the design of routers and switches. For example, our results suggest that core Internet routers receive traffic that is much less self-similar than traffic that emerges out of border routers directly connected to Ethernet LANs.

#### Acknowledgments

The simulation code to generate the self-similar traffic was partly written by Haiguang Cheng. The authors would also like to gratefully acknowledge his contribution through discussions that further improved the traffic model.

#### REFERENCES

- [1] S. Borst, O. Boxma and P. Jelenković. "Asymptotic Behavior of Generalized Processor Sharing with Long-Tailed Traffic Sources". *Proceedings of IEEE INFOCOMM*, Tel Aviv, Israel, Mar. 2000.
- [2] M. E. Crovella and A. Bestavros. "Self-Similarity in World Wide Web Traffic: Evidence and Possible Causes". *IEEE/ACM Transactions on Networking*, vol 5, no. 6, Dec. 1997.
- [3] S. Fong and S. Singh. "Performance Evaluation of Shared-Buffer ATM Switches Under Self-Similar Traffic". *Proceedings of IEEE International Performance, Computing, and Communications Conference*, Arizona, 1997.
- [4] R. G. Garroppo, S. Giordano, S. Miduri, M. Pagano and F. Russo. "Statistical Multiplexing of Self-Similar VBR Video-conferencing Traffic". *Proceedings of GLOBECOM*, 1997.
- [5] A. Gersht, G. Pathak and A. Shulman. "Burst Level Congestion Control in ATM Networks". *Proceedings of IEEE Symposium on Computer and Communications*, Athens, Greece, July 1998.
- [6] W. E. Leland, M. S. Taqqu, W. Willinger and D. V. Wilson. "On the Self-Similar Nature of Ethernet Traffic (Extended Version)". *IEEE/ACM Transactions on Networking*, vol. 2, no. 1, Feb. 1994.
- [7] G. Samorodnitsky and M. S. Taqqu. *Stable Non-Gaussian Random Processes: Stochastic Models with Infinite Variance*. Chapman & Hall, NY, 1994.
- [8] S. Song, J. K.-Y. Ng and B. Tang. "On the Self-Similarity Property of the Output Process from a Network Server with Self-Similar Input Traffic". *Proceedings of Sixth International Conference on Real-Time Computing Systems and Applications*, pp. 128-132, Dec. 1999.
- [9] W. Stallings. *High-Speed Networks: TCP/IP and ATM Design Principles*. Prentice Hall, Upper Saddle River, NJ, 1998.
- [10] B. Tsybakov and N. D. Georganas. "Overflow Probability in an ATM Queue with Self-Similar Input Traffic". *Proceedings of IEEE International Conference on Communications*, Montreal, Canada, June 1997.
- [11] W. Willinger, M. S. Taqqu, R. Sherman and D. V. Wilson. "Self-Similarity Through High-Variability: Statistical Analysis of Ethernet LAN Traffic at the Source Level". *IEEE/ACM Transactions on Networking*, vol.5, no.1, Feb. 1997.
- [12] S. Wittevrongel and H. Bruneel. "Output Traffic Analysis of a Leaky Bucket Traffic Shaper Fed by a Bursty Source". *Proceedings of IEEE International Conference on Communications*, pp. 1581-1585, 1994.
- [13] X. Yang, A. P. Petropulu and V. Adams. "The Extended ON/OFF Model for High-Speed Data Network". *Workshop on Applications of Heavy Tailed Distributions in Economics, Engineering and Statistics*, Washington, D.C., Jun. 1999.
- [14] H. Zhang. "Service Disciplines for Guaranteed Performance Service in Packet-Switching Networks." *Proceedings of the IEEE*, vol. 83, no. 10, Oct. 1995.

# PARAMETER ESTIMATION IN FARIMA PROCESSES WITH APPLICATIONS TO NETWORK TRAFFIC MODELING

Jacek Ilow

Dalhousie University  
Department of Electrical and Computer Engineering  
Halifax, NS, B3J 2X4, Canada  
e-mail: j.ilow@dal.ca

## ABSTRACT

Traffic measurements in many network environments demonstrate the coexistence of both long- and short-range dependence in traffic traces. In this paper, we use the fractionally integrated autoregressive moving average (FARIMA) processes with non-Gaussian innovations to describe packet arrival rate in a unit time. Specifically, we investigate cepstrum-based approaches for parameter estimation in FARIMA processes. We examine the fractional differencing parameter estimation procedure based on the smoothed periodogram and the log spectrum. The simulation results demonstrate that the proposed cepstrum approach gives better estimation accuracy than the conventional least-square spectrum fit. Usefulness of the results presented is demonstrated on the real network traffic traces by considering spectral fitting metrics.

## 1. INTRODUCTION

The objective of traffic characterization is to transform complicated intrinsic processes in the network into a nearly equivalent traffic model which is credible, analytically tractable and computationally efficient. Traffic modeling is important in many areas of network engineering such as design, control and performance evaluation. Conventional models for the network traffic include: pure Poisson and Markov-Modulated Poisson processes; packet-train; fluid flow and autoregressive-moving average (ARMA) models. However, it has been argued that these models do not represent completely the long-range dependence (LRD) property of the network traffic discovered recently by researchers from Bellcore [1].

Long memory processes are able to capture a slowly decaying auto-correlation structure of the underlying time series [2]. Suitable approaches for the generation and representation of long memory processes include fractional Gaussian noise, fractionally integrated ARMA processes and chaotic maps. The fractional Gaussian noise model has been used to account for LRD in the Ethernet traffic data in most of the publications. However, as we will show in this paper, network traffic is usually not Gaussian distributed and its autocorrelation cannot be represented accurately by the single-parameter (the Hurst parameter) as is the fractional Gaussian noise model. The FARIMA model generalizes

the broad class of ARMA processes, and includes the fractional differencing (strictly self-similar) model as a special case. As such, FARIMA modeling encompasses and enriches currently used models [3].

There are two methods one can adopt for model building: the first relies on a theoretical model formulation from basic events in the network; the second employs experimental data fitting. In this paper, we take the second approach. The model validation is achieved by employing statistical tests of goodness-of-fit for the dependence in time series [4]. Since the traffic data is usually non-Gaussian, we propose to use the polyspectra approach [5] to estimate the parameters of an FARIMA model. This approach does not make any assumption on the marginal distribution of data except that the second and third order moments are finite. In addition, it does not require a priori knowledge of the order of the ARMA part and can identify non-minimum phase systems.

The data sets used in the experimental section of this paper are part of a large number of high-resolution Ethernet measurements recorded by Bellcore, Morristown. We analyze traces obtained from the URL site [6] which are pre-processed to give traffic workload (i.e., a number of bytes in a unit time).

## 2. PRELIMINARIES OF FARIMA PROCESSES

An FARIMA process  $Y_t$  with parameters  $(n, d, m)$  is defined through the following difference equation [2]:

$$\Phi(z^{-1}) \cdot (1 - z^{-1})^d Y_t = \Theta(z^{-1}) \epsilon_t, \quad (1)$$

where  $t$  indicates discrete time;  $Y_t$  is the observed time sequence;  $\epsilon_t$  is an i.i.d. non-Gaussian sequence with finite mean and variance;  $z^{-1}$  is the back-shift operator;  $\Phi(z^{-1})$ , and  $\Theta(z^{-1})$  are the autoregressive (AR) and moving average (MA) polynomials of order  $m$  and  $n$ , respectively. The fractional differencing is defined by the binomial series expansion [2]:

$$(1 - z^{-1})^{-d} \triangleq \sum_{j=0}^{\infty} \alpha_j z^{-j}, \quad (2)$$

where the coefficients  $\alpha_j$  are given through the recursive formula:

$$\alpha_0 = 1, \quad \alpha_j = \alpha_{j-1} \frac{j-1+d}{j}. \quad (3)$$

The FARIMA process  $Y_t$ , defined in (1), can be interpreted as the output of the system (Fig. 1) driven by  $\epsilon_t$  with the transfer function  $H(z^{-1}) = \frac{1}{(1-z^{-1})^d} H_{ARMA}(z^{-1})$ , where  $H_{ARMA}(z^{-1}) = \frac{\Theta(z^{-1})}{\Phi(z^{-1})} = \sum_{i=0}^{\infty} h_i z^{-i}$  is the ARMA part of the system, and  $\{h_i\}$  is its impulse response.

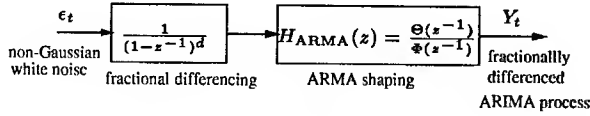


Figure 1: Model of a fractionally differenced ARIMA (FARIMA) process.

### 3. IDENTIFICATION OF AN FARIMA MODEL

In this section, we describe a two-step parameter estimation procedure for FARIMA processes based on the available observations  $\{Y_t, t = 1, \dots, N\}$ . First, we obtain an estimate  $\hat{d}$  of the parameter  $d$  based on the log of the power spectrum. This is carried out independently of the ARMA part of the model. Second, we estimate the impulse response of the ARMA part of the system using the polycepstral approach [5]. In the ARMA estimation, we operate on data  $\{X_t\}$  obtained by passing  $\{Y_t\}$  through  $(1-z^{-1})^d$ . The two-step estimation scheme described in this section is illustrated in Fig. 2.

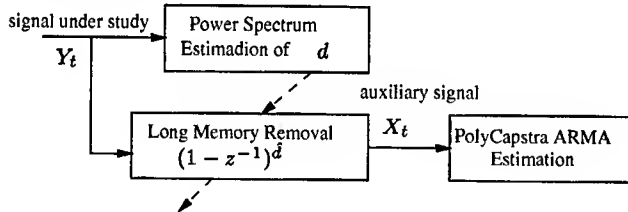


Figure 2: Proposed algorithm for FARIMA processes parameter estimation.

#### 3.1. Estimation of the Fractional Differencing Parameter

To estimate the fractional differencing parameter  $d$ , we adopted from [7] a technique based on the log of power spectrum or, equivalently, power cepstrum. With the "spectrum" of  $\{Y_t\}$  given as:

$$f_Y(\omega) = |1 - \exp^{-j\omega}|^{-2d} f_{ARMA}(\omega), \quad (4)$$

where  $f_{ARMA}(\omega)$  is the spectrum shaping from the ARMA filtering, the log of the power spectrum can be expressed in

terms of the cepstral coefficients  $\{c_k\}$  as follows [7]:

$$\log f_Y(\omega) = \sum_{k=1}^{\infty} c_k \cos(k\omega), \quad (5)$$

where

$$c_k = \frac{1}{\pi} \int_0^{\pi} \log[f_Y(\omega)] \cos(k\omega) d\omega. \quad (6)$$

Using the weight function  $W(\omega) = -0.5 \log[2(1 - \cos \omega)]$ , we define the weighted power cepstrum index  $S$  as:

$$S \triangleq \frac{1}{\pi} \int_0^{\pi} W(\theta) \log f_Y(\theta) d\theta. \quad (7)$$

With this, it can be shown that [7]:

$$S = d \sum_{k=1}^{\infty} \frac{1}{k^2} + \sum_{k=1}^{\infty} \frac{a_k}{2k}, \quad (8)$$

where  $a_k$  are power cepstrum coefficients of the ARMA part of the spectrum.

Because  $a_k$  decays exponentially as  $k$  increases [5], we will assume that above certain threshold value  $M$ ,  $a_k = 0$ , for  $k > M$ . Then, by estimating the weighted power cepstrum index  $S$  and coefficients  $c_k$  up to  $k < M$ , we obtain  $\hat{d}$ , based on (8), in the following way:

$$\hat{d} = \frac{1}{\frac{\pi^2}{6} - \sum_{k=0}^{M-1} \frac{1}{k^2}} (\hat{S} - \sum_{k=0}^{M-1} \frac{1}{k} \hat{c}_k), \quad (9)$$

where  $\hat{c}_k$  and  $\hat{S}$  are estimates of  $c_k$  and  $S$ , respectively. To obtain  $\hat{c}_k$  and  $\hat{S}$ , we use the periodogram  $I_N(\omega)$  evaluated based on  $N$  data points. With the Simpson rule for calculating integrals,  $c_k$  is estimated from (6) as:

$$\hat{c}_k = \frac{1}{\pi} \sum_{p=0}^{\{N/2\}} \log I_N(\omega_p) \cos(k\omega_p). \quad (10)$$

The estimate  $\hat{S}$  is obtained based on (7), in a similar way as  $\hat{c}_k$  in (10).

An alternative approach to calculate the  $d$  parameter using the least-squares fit of the FD model to the periodogram has been presented in [8].

#### 3.2. Estimation of ARMA parameters

Instead of finding the AR and MA polynomials,  $\Phi(z^{-1})$  and  $\Theta(z^{-1})$ , we employ the estimation procedure for the impulse response  $\{h_i\}$  of the ARMA filter  $H_{ARMA}(z^{-1})$  based on the new observable data  $\{X_t, t = 1, \dots, N\}$ , as shown in Fig. 2. We assume that the transfer function  $H_{ARMA}(z^{-1})$  admits the factorization:

$$H_{ARMA}(z^{-1}) = A \cdot I(z^{-1}) \cdot O(z), \quad (11)$$

where  $A$  is a constant gain;  $I(z^{-1})$  and  $O(z)$  are a minimum phase and a maximum phase polynomials. The impulse response  $\{h_i\}$  is obtained in the following way: first we calculate the unknown impulse responses  $\{i_k\}$  and  $\{o_k\}$  of the minimum phase and maximum phase characteristics of

the system, and then we obtain  $\{h_k\}$  as a convolution of  $\{i_k\}$  and  $\{o_k\}$  ( $h = i * o$ ).

The estimation procedure employed use the minimum and maximum phase differential cepstrum coefficients which, for non-symmetric data  $X_t$ , can be calculated from two slices of the bicepstrum as described in [5].

#### 4. MODEL VALIDATION

##### 4.1. Non-Gaussianity due to skewness and kurtosis

When estimating the ARMA part of the FARIMA model in the previous section, we made an assumption that the observed process  $\{Y_t\}$  was (i) non-Gaussian and (ii) asymmetric. One can single out two distinct deviations of histogram from the Gaussian distribution:

- one of the tails of the distribution is lengthened, and the distribution becomes skewed (asymmetric);
- the maximum of the histogram (pdf) lies higher or lower than that of the normal distribution.

The shape statistics that characterize these deviations are standardized third and fourth order moments which for a Gaussian population are  $b = 0$  and  $a = 3$ , respectively. Departures from these values are indication of skewness and kurtosis, respectively. To reject the hypothesis due to skewness at the 95% significance level that the data of length 2000 are Gaussian, it is sufficient to show that  $b > 0.09$  or  $b < -0.09$ . The same data fail the test for Gaussianity due to kurtosis if  $a > 3.18$  or  $a < 2.83$ .

##### 4.2. Goodness-of-fit test for the dependence structure

In this paper, we analyze a model according to the modified version of the portmanteau lack of fit test [4]. First, we compute the following test statistics:

$$T \triangleq \frac{\frac{N}{4\pi} \sum_{i=0}^{N-1} \left( \frac{f_{\Theta}(\omega_i)}{I_N(\omega_i)} \right)^2}{\left( \sum_{i=0}^{N-1} \frac{f_{\Theta}(\omega_i)}{I_N(\omega_i)} \right)^2}, \quad (12)$$

where  $f_{\Theta}(\cdot)$  is the spectral density of the model parameterized by  $\Theta$ . The test statistics, as defined in (12), measures departures of the modeled spectral density from the periodogram in the whole range of frequencies  $(-\pi, \pi]$ . In the time domain,  $T$  is the sum of the squares of all estimable correlations of the residual process obtained by fitting the chosen model, and as such is especially useful for long memory processes [2]. If the parametric model is correct, then the residual process is uncorrelated, and  $T$  should be close to 0. The test statistics  $T$  is asymptotically normal, and it can be shown that  $Pr(T < c) \simeq \Phi(\sqrt{N}(c\pi - 1)/\sqrt{2})$ , where  $\Phi(\cdot)$  is the cumulative distribution function (cdf) of the standard normal RV. In this paper, we use as a discriminating measure between different models the P-value of the test statistics in (12) defined as  $Pr(T < T^*)$ , where  $T^*$  is the outcome of the test statistics for a given data set. A large P-value indicates that we have a correct model, while a small value supports the hypothesis that the model is inaccurate.

#### 5. SIMULATION AND EXPERIMENTAL RESULTS

In this section, we first examine the performance of the estimation procedure proposed using simulated data, and then we demonstrate the effectiveness of the FARIMA model for network traffic.

##### 5.1. Simulated Data

Table 1 presents estimation results using the simulated FARIMA data. Three different types of the ARMA part are considered: (i) AR with  $H_{ARMA}(z) = \frac{1}{1-0.5z^{-1}}$ ; (ii) MA with  $H_{ARMA}(z) = 1-0.5z^{-1}$ ; and (iii) ARMA with  $H_{ARMA}(z) = \frac{1-0.2z^{-1}}{1-0.5z^{-1}}$ . We give the average and standard deviation values (in parentheses) of Monte-Carlo simulation results based on processing 50 independent blocks of data; each of them with  $2^{12}$  samples. The system driving noise was zero-mean, white, non-Gaussian (exponentially distributed). While estimating  $d$ , we used the value of  $M = 15$  beyond which we assumed that the power cepstral coefficients from the ARMA part are not significant. It can be observed that the variance of the  $\hat{d}$  estimator increases as  $d$  goes from 0.1 to 0.4. The  $\hat{d}$  estimator is biased [7], but in general, a good fit was observed to the model when the parameters  $n$  and  $m$  were small ( $n, m \leq 5$ ).

The estimation method for the  $d$  parameter used in this paper gives much better results than the least-squares method presented in [8].

In Fig. 3, we present the P-values for FAR (FARIMA(0,d,m)), FMA (FARIMA(n,d,0)) and AR models as a function of the AR or MA approximation orders. The results are for data which were generated by passing one-sided exponential noise through the filter  $(1-z^{-1})^{-0.4} \cdot (1-0.5z^{-1})$ . Each point in this figure is the average of P-values calculated for 20 blocks of data of length  $2^{12}$ . As we see, the P-value can indicate the correct ( $n \sim 1$ ) order of the FMA representation and shows that no better fit is obtained by using higher orders of an MA part.

##### 5.2. Ethernet Data

In this section, we apply the FARIMA model to the traffic traces which were obtained from the URL site [6]: BC-pOct89 and BC-Oct89Ext. The first trace represents *internal traffic* on the Bellcore LAN, while the second trace represents *external traffic* from Bellcore to the outside Internet world.

##### Workload of the Internal Traffic

The trace BC-pOct89 contains traffic for about 30 min ( $10^6$  Ethernet packets in 1,759 sec). This data sets was first pre-processed into time series to give the number of bytes in 10 millisecond intervals. In our analysis, we considered only 20 blocks, each of  $2^{12}$  samples, which gives a rise to 13.2 minutes of traffic. In such a time interval, we can assume that the internal traffic environment is stationary [7]. For each block of data, we fitted four models: (i) FAR; (ii) FMA; (iii) AR of order 10 through least-squares; and (iv) fractional differencing (FD) model. For the particular



Table 1: Statistical behavior of the proposed parameter estimation method for FARIMA processes.

$HARMA(z^{-1})$	$\frac{1}{1-0.5z^{-1}}$ $\theta_1 = -0.5$		$1 - 0.5z^{-1}$ $\phi_1 = -0.5$		$\frac{1-0.2z^{-1}}{1-0.5z^{-1}}$ $h_1 = -0.3, h_2 = -0.06, h_3 = -0.012$			
$d$	$\hat{d}$	$-\hat{\theta}_1$	$\hat{d}$	$-\hat{\phi}_1$	$\hat{d}$	$-\hat{h}_1$	$-\hat{h}_2$	$-\hat{h}_3$
0.1	.111 (.6e-2)	0.51 (1.0e-2)	.109 (1.9e-2)	0.497 (1.4e-2)	.12 (1.9e-2)	.3 (1.4e-3)	.05 (1.8e-3)	.03 (9.0e-3)
0.2	.217 (7.1e-2)	0.52 (1.2e-2)	.219 (8.7e-2)	0.508 (4.1e-2)	.195 (8.2e-2)	.348 (4.8e-3)	.087 (1.9e-3)	.009 (4.2e-3)
0.3	.327 (1.03e-1)	0.511 (6.2e-2)	.319 (1.21e-1)	0.538 (2.6e-1)	.326 (1.23e-1)	.228 (4.6e-3)	.0567 (3.15e-3)	.03 (4.2e-3)
0.4	.455 (1.01e-1)	.541 (6.2e-2)	.413 (1.21e-1)	.44 (2.6e-2)	.428 (1.23e-1)	.4 (7.6e-3)	.0967 (5.15e-3)	.014 (2.2e-3)

trace, the averaged indexes of skewness and kurtosis from 20 blocks of data of length 2000 were as follows:  $b = 0.4$ ,  $a = 2.8$ . This indicates that these data are non-Gaussian and non-symmetric.

To assess goodness-of-fit of each model into the dependence structure of the underlying time series, we first examine visually the fit of the models (their transfer functions) to the power spectral density (PSD) of the trace. The PSD is estimated by ensemble averaging of periodograms evaluated in each block. The fits of the estimated models to the averaged periodogram are shown in Fig. 4. We present results using the log-log and dB scale. The log-log plot emphasizes the low frequency region, while dB plot gives an idea about the overall fit. It is evident that the FMA model with just 3 and 4 coefficients in the MA part offers the best fit. The estimated fractional differencing parameter is 0.3395. To capture the short-range dependence of the trace (or to obtain a better fit in the high frequency region), we used a fourth order MA representation ( $h = \{1, -0.40, -0.17, 0.04, 0.14\}$ ).

The P-values of the estimated models are shown in Fig. 5. This test confirms our intuitive observations based on the periodogram analysis. Because the P-value measures the overall fit of a model to the periodogram, the performance of the FD model is worse than that of the AR model of order greater than 5. It is evident that FMA and FAR models give the most parsimonious representations.

### Workload of the Eternal Traffic

The trace BC-Oct89Ext represents around 34 hours of external traffic. We pre-processed the data to get the number of bytes in 1 second intervals. We apply the same analysis to BC-Oct89Ext as for the internal traffic. The averaged indexes of skewness and kurtosis is 2.3 and 4.2, respectively. This shows that the external traffic is also non-Gaussian and non-symmetric. The PSD based on the periodogram and three fitted models and the P-values of the modified portmanteau test statistics are shown in in Figs. 6 (a) and (b). Apparently, the external traffic is fitted well by the FD model. There is no significant improvement by using the FARIMA approach. The fractional differencing parameter in this case is  $\hat{d} = 0.3981$ , which indicates heavy burstiness of the trace.

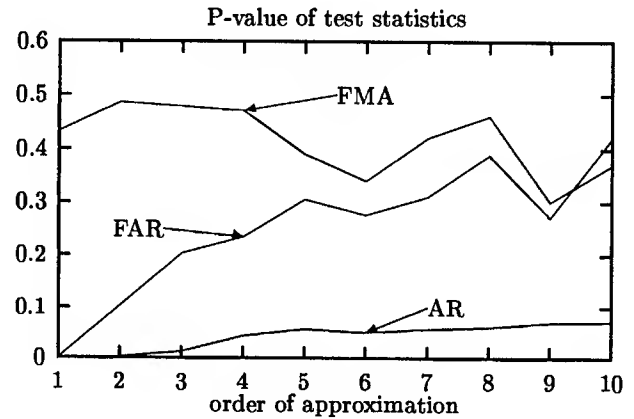


Figure 3: P-value of the modified portmanteau test statistics for three types of fitted models to the simulated data with  $H(z) = [(1 - z^{-1})^{-0.4}] (1 - 0.5z^{-1})$ : (i) FMA; (ii) FAR; and (iii) AR.

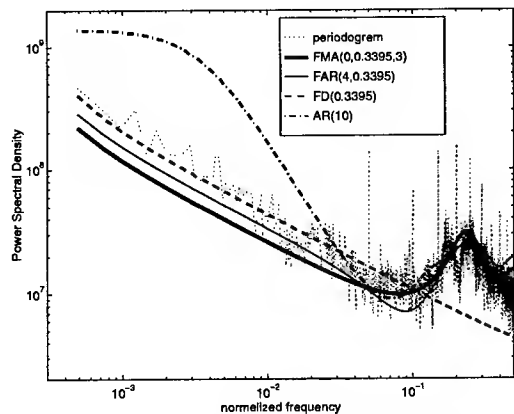
### 6. CONCLUSION

In this paper, we extended the long memory modeling to include the short-term dependence in the data by using FARIMA processes. Because of non-Gaussianity of network traffic, we developed a two stage parameter estimation scheme for the FARIMA model using the polyspectra approach. We evaluated the effectiveness of the proposed model using real network traffic data. The following observations are made: (i) the model proposed provides a better fit to the internal LAN traffic than the conventional least-squares AR and fractional differencing models; and (ii) the external LAN traffic is well characterized by fractionally differenced model. In conclusion, the proposed method can capture the complex dependence structure in network traffic with a small number of parameters.

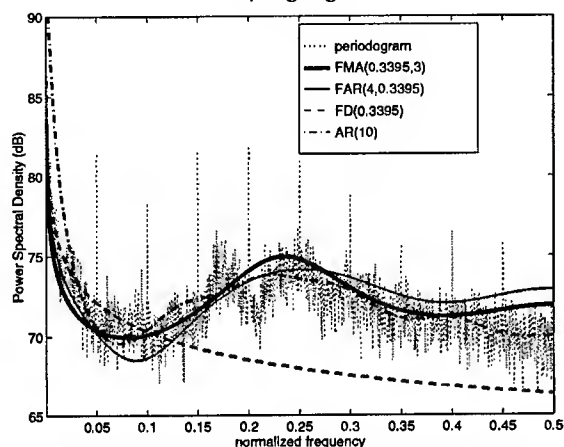
### REFERENCES

- [1] W. Leland, M.S. Taqqu, W. Willinger and V. Wilson, "On the self-similar nature of ethernet traffic (extended version)," *IEEE/ACM Trans. Networking*, vol. 2, No.1, pp. 1-15, Feb. 1994.
- [2] J. Beran, *Statistics for Long-memory Processes*. Chapman & Hall, 1994.



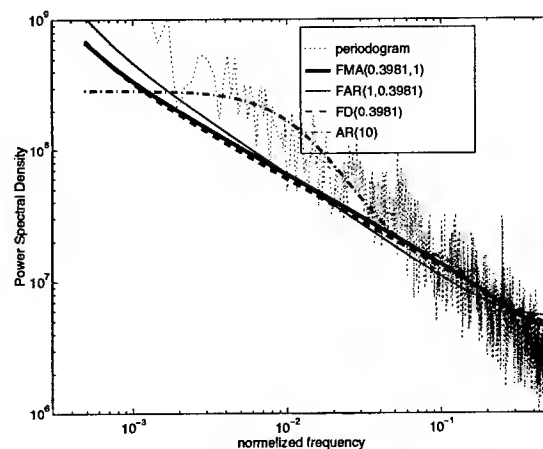


a) log-log

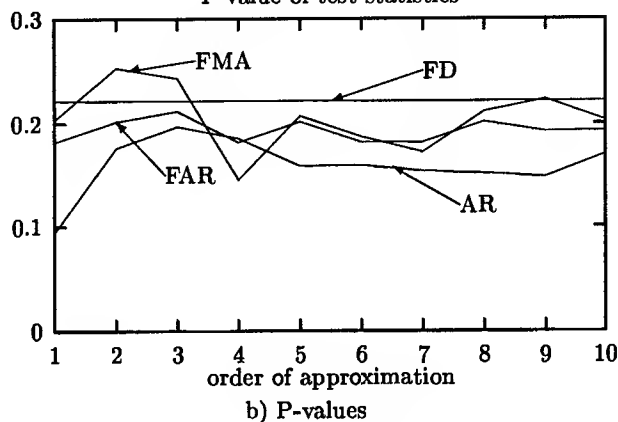


b) dB plot

Figure 4: Power Spectral Density for the BC-pOct89 trace of Ethernet traffic based on periodogram and four fitted models: (i) FMA; (ii) FAR; (iii) FD; and (iv) the least-square AR(10).



a) PSD (log-log)  
P-value of test statistics



b) P-values

Figure 6: External traffic trace BC-Oct89Ext: (a) Power Spectral Density based on periodogram and four fitted models: (i) FMA; (ii) FAR; (iii) FD; and (iv) the least-squares AR(10); (b) P-values of the fitted models.

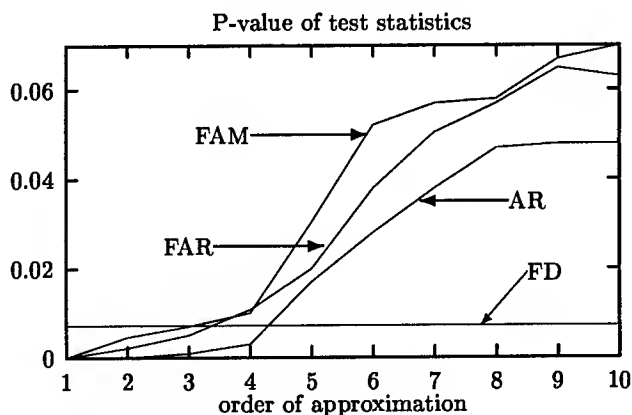


Figure 5: P-values of the modified portmanteau test statistics for four types of fitted models for the trace BC-pOct89.

- [3] J. Hosking, "Fractional differencing," *Biometrika*, vol. 68, No.1, pp. 165-176, 1981.
- [4] J. Beran, "A goodness-of-fit test for time series with long range dependence," *J.R. Statist. Soc. B*, vol. 54, No.3, pp. 749-760, 1992.
- [5] C. Nikias and A. Petropulu, *Higher-Order Spectra Analysis: A Nonlinear Signal Processing Framework*. Prentice Hall Signal Processing Series, 1993.
- [6] URL address <http://town.hall.org/Archives/pub/ITA>.
- [7] G. Janecek, "Determining the degree of differencing for time series via the log spectrum," *Journal of Time Series Analysis*, vol. 3, No.3, pp. 177-183, March 1982.
- [8] J. Ilow, "Forecasting network traffic using FARIMA models with heavy tailed innovations," *ICASSP'00, Istanbul, Turkey*, June 2000.

# Nonlinear Filtering Algorithm with its application in INS Alignment

Rui Zhao , Qitai Gu

Department of Precision Instruments and Mechanology,  
Tsinghua University Beijing, China, 100084  
E-mail: [zhaorui@post.pim.tsinghua.edu.cn](mailto:zhaorui@post.pim.tsinghua.edu.cn)

## ABSTRACT

The application of optimal nonlinear/non-Gaussian filtering to the problem of inertial navigation system (INS) alignment is described. This approach is made possible by a new technique called particle filtering (PF). PF theory is introduced and nonlinear error equations of INS alignment on a stationary base in the case of large initial error angles are used. The algorithm for solving the problem of optimal estimation of the state vector described by nonlinear equations from linear measurements has been developed. The simulation results exhibit the superior performance of this approach when compared with classical sub-optimal techniques such as extended Kalman filtering (EKF).

## 1.INTRODUCTION

Kalman filtering is a popular tool in handling estimation problems, but its optimality heavily depends on linearity. When used for nonlinear systems, its performance relies on, and is limited by the linearizations performed on the concerned model. For those essential nonlinear systems, the linearizations may lead to divergence of filtering process. On the other hand, despite early papers on nonlinear filtering theory, the implementation of nonlinear filters has been plagued so far by the difficulties inherent to their infinite-dimensional nature. A new approach to optimal nonlinear filtering called particle filtering (PF) has been

presented recently, which is applied to the Non-Gaussian/Nonlinear filtering problem [1][2][3][4]. The main feature of PF is that it constructs the conditional probability of the variable to be estimated, with respect to the measurements, through a suitable random particle exploration of the state space followed by a Bayes correction of the weights of the particles.

## 2.THE THEORY AND PRIORI

### ALGORITHM OF PARTICLE FILTERING

Let the dynamic process  $X$  and the observation process  $Y$  be governed by

$$\begin{cases} X_{k+1} = f(X_k, k, \omega_k) \\ Y_k = h(X_k, k) + \eta_k \end{cases}$$

where  $\{\omega_k\}$  and  $\{\eta_k\}$ ,  $k \geq 0$ , are sequences of independent random variables with appropriate dimensions.  $R^n$  is defined as the strength matrix of  $\eta_k$ , which is assumed to be strictly positive definite.  $f$  and  $h$  are measurable functions of  $X$ . PF concerns the recursive estimation of any function  $\Phi(X_k)$  of an  $R^n$ -valued stochastic process  $X$  from the observation of a related  $R^m$ -valued, random process  $Y$ , where the "best" (minimum variance) estimator  $\Phi(X_k)^*$  is given by the conditional expectation

$$E[\Phi(X_k) | Y_K = y_K] = \int_{R^n} \Phi(x_k) dP(x_k | y_K)$$

(with notation  $K$  stands for the sequence  $1, 2, \dots, k$ )

The priori PF algorithm may be summarized as follows[5] :

a) Initialization.

Positions of  $N$  particles are initialized according to  $dP(X_0)$  and the weights to  $1/N$ .

b) Evolution.

Move particles according to  $X_{k+1} = f(X_k, k, \omega_k)$  and randomly generated noises  $\omega_k$ .

c) Weighting.

Weights are given by

$$W_k^i = \frac{Z_k^i}{\sum_{j=1}^N Z_k^j}$$

and regularization according to ( in the case of a Gaussian observation process )

$$Z_k^j(X, Y) = \frac{\exp\left\{-\frac{1}{2} \sum_{i=1}^k \gamma^{k-i} \|Y_i - h(X_i^j, i)\|_{R^n}^2\right\}}{\exp\left\{-\frac{1}{2} \sum_{i=1}^k \gamma^{k-i} \|Y_i\|_{R^n}^2\right\}}$$

here  $\gamma \in (0, 1)$

(with notation:  $\|a\|_{R^n}^2 = a^T [R^n]^{-1} a, a \in R^n$ )

d) Estimation.

According to

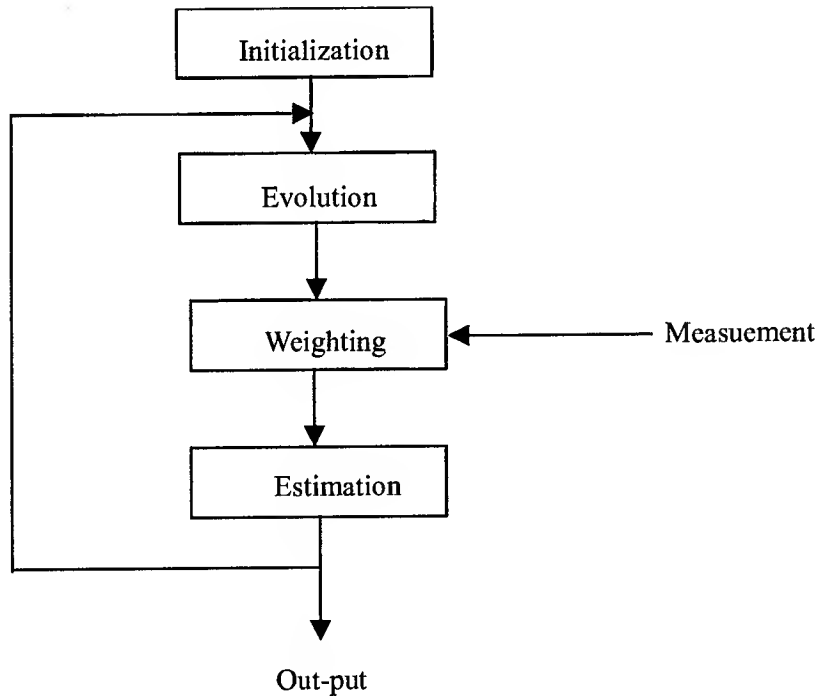
$$\phi(x_k)^* = \sum_{i=1}^N W_k^i \phi(x_k^i)$$

the estimation of  $\Phi(X_k)$  is made.

e) Recursion.

Step from b to d.

Figure 1 depicts the procedure of the priori algorithm.



**Figure 1.** Block diagram of the priori PF algorithm.

### 3. THE APPLICATION OF PF IN INS ALIGNMENT

The problem of INS alignment has been considered in a great number of publications[5][6]. In the case of small initial error angles, the problem is often solved on the basis of a linearized description of INS error equations and an elaborate procedure of the optimal linear filtering proceeding from the condition to achieve the maximum accuracy over the minimum time. At the same time, in a number of cases INS is often to be aligned under the conditions that the initial error angles are comparably large which makes it necessary to take account of nonlinear character of the problem. However, no detailed study of the alignment problem with due account of its nonlinear character has been made up to now. This paper gives a treatment of INS alignment problem by PF. The Global Positioning Systems (GPS) are to be used as an external measuring instrument for the information about the carrier position. For a stationary carrier the information of its zero velocity and acceleration is applied. On the basis of Dmitriyev's work[6], nonlinear error equations of INS alignment on a stationary base in the case of large initial error angles are as follows:

$$\delta \dot{v}_x^l = -g(\phi_y \cos \phi_z + \phi_x \sin \phi_z) + 2\omega_{ie} \sin L \delta v_y + \nabla_x$$

$$\delta \dot{v}_y^l = g(\phi_x \cos \phi_z - \phi_y \sin \phi_z) - 2\omega_{ie} \sin L \delta v_x + \nabla_y$$

$$\dot{\phi}_x^l = -\sin \phi_z \omega_{ie} \cos L + \phi_y \omega_{ie} \sin L - \delta v_y / R - \varepsilon_x$$

$$\dot{\phi}_y^l = (1 - \cos \phi_z) \omega_{ie} \cos L - \phi_x \omega_{ie} \sin L + \delta v_x / R - \varepsilon_y$$

$$\dot{\phi}_z^l = (\phi_x \cos \phi_z - \phi_y \sin \phi_z) \omega_{ie} \cos L + \delta v_x (\tan L) / R - \varepsilon_z$$

and the observation equation may be written as

$$y_1 = \delta v_x + \eta_x$$

$$y_2 = \delta v_y + \eta_y$$

The nonlinear equations describe the behavior of the INS alignment errors exactly. With the scope of PF theory, the INS alignment is formulated as the problem of optimum estimation of error angles described by means of nonlinear equations from linear measurements. The validation of this method was checked by simulation as follows. The priori algorithm with N=1500 particles was used in the simulation. Figure 2 Shows the filters outputs, i.e., the RMS deviation of the error angles  $\Phi_x$  and  $\Phi_z$ , as estimated by the PF(solid line) and EKF(dashed line).

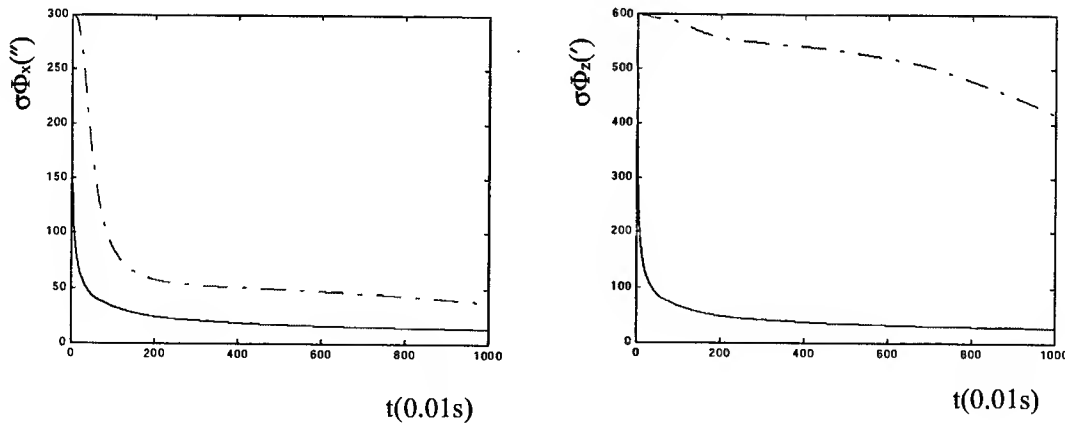


Figure 2. The RMS deviation of estimation.

## 4. CONCLUSION

These results show the clear superiority of particle nonlinear filtering over those classical filtering in the problem of INS alignment in the

case of big initial error angles, although the former is more time/memory consuming as the number of particles grows. These problems are overcome by the advent of new technologies, making parallel processing available to embedded systems, and enabling PF to be implemented in on-board real-time systems.

From a long run, PF is sure to be a powerful tool dealing with nonlinear/non-Gaussian filtering problems such as INS alignment.

## 5. REFERENCES

- [1] Huillet T., and Salut G. "Interprétation des équations du filtrage non-linéaire". In *Séances du GDR Automatique du CNRS (Pôle non-linéaire)*, Paris, Nov, 1989.
- [2] Carvalho H., Monin A., and Sault G. "Filtrage optimal non-linéaire du signal GPS NAVSTAR en recalage de centrales de navigation". *Report LAAS*, July, 1994.
- [3] Del Moral P. "Résolution particulière des problèmes d'estimation et d'optimisation non-linéaires". *Thèse de l'Université Paul Sabatier, LAAS-CNRS*, Toulouse, June 1994.
- [4] Gordon N. J., Salmond D. J., and Smith A. F. M. "Novel approach to nonlinear/non-Gaussian Bayesian state estimation". *IEEE Proceedings*, Pt. F, Oct 1992, pages 107-113.
- [5] Carvalho H., Moral P. "Optimal Nonlinear Filtering in GPS/INS Integration". *IEEE Trans. AES*, 33(3):835-849, 1997.
- [6] Dmitriyev S. P., Stepanov O. A. "Nonlinear filtering methods application in INS alignment". *IEEE Tran. AES*, 33(1):260-271, 1997.

# GPS JAMMER SUPPRESSION WITH LOW-SAMPLE SUPPORT USING REDUCED-RANK POWER MINIMIZATION\*

Wilbur L. Myrick & Michael D. Zoltowski

Purdue University  
School of Electrical Engineering  
1285 Electrical Engineering Bldg.  
West Lafayette, IN 47907-1285  
[wlm, mikedz]@ecn.purdue.edu

J. Scott Goldstein

SAIC  
Adaptive Signal Exploitation  
4001 N. Fairfax Drive  
Arlington, VA 22203  
sgoldstein@trg1.saic.com

## ABSTRACT

When wideband and narrowband interferences in a GPS system are stationary, a large number of data samples may be obtained to get a good estimate of the interference. However, the jamming environment may be one in which the narrowband jammers have the ability to change frequencies dynamically or the rapid dynamics of the aircraft during maneuvering causes arrival angles of wideband jammers to change. In either type of jamming environment, an interference suppression algorithm will only be effective if it can rapidly converge with a small sample size. We investigate the performance of reduced-rank interference suppression algorithms under conditions of low sample support. It is demonstrated that the multistage nested Wiener filter (MSNWF) outperforms other reduced-rank techniques in terms of suppressing both wideband and narrowband jammers under conditions of low sample support.

## 1. INTRODUCTION

Worldwide military use of GPS is evolving due to the wide availability of commercial GPS receivers, and the widespread knowledge of the force enhancement capabilities offered by GPS. The jamming threat is serious because of the physical design of the GPS system. The received power from the GPS satellites is approximately -157 dBW. Many jammers available on the arms market today either already cover the GPS frequencies, or can be modified to do so. Therefore, a space-time preprocessing filter prior to the GPS correlators is one of several proposed methods for suppressing such jammers. However, space-time preprocessors can exhibit slow convergence and have high computational complexity.

This paper investigates a reduced dimension space-time preprocessor based on the multistage nested Wiener filter (MSNWF)[4] capable of operating with a low sample support compared to other reduced dimension methods such as cross-spectral method[3] and principal components. The simulations presented herein reveal the rapid convergence of the MSNWF implementation of the power minimization based space-time preprocessor, thereby showing its efficacy

THIS RESEARCH WAS SUPPORTED BY THE AIR FORCE OFFICE OF SCIENTIFIC RESEARCH UNDER CONTRACT NO. F49620-97-1-0275.

in adapting to the environmental dynamics characterizing high performance fighter aircraft.

## 2. POWER MINIMIZATION BASED JOINT SPACE-TIME PREPROCESSOR

The criterion for determining the optimal set of space-time weights is premised on the fact that the respective power levels of the desired GPS signals are significantly below the noise floor and that the jammers that could have deleterious effects are above the noise floor. The goal then is to drive the power of the preprocessor output down to the noise floor. This approach serves to place point nulls at the respective angle-frequency coordinates of strong narrowband interferers and spatial nulls in the respective directions of broadband interferers.

In order for the GPS receiver to provide accurate navigation information, it is necessary to track the signals from at least four different GPS satellites. Given the parallax error associated with GPS satellites at near-horizon relative to the aircraft, it is generally desirable to track the respective signals from a larger number of GPS satellites, e.g., twelve. It is desired then that the preprocessor "pass" unaltered as many GPS signals as possible. Thus, the magnitude of the multidimensional Fourier transform of the space-time weights should be as flat (smooth) as possible in the spectral domain as a function of frequency and angular dimensions. The goal is to achieve a desired smoothness while simultaneously nulling both wideband and narrowband interferers under conditions of low sample support.

### 2.1. Formulation of Objective Function

It is necessary to first define  $\mathbf{x}_m(n)$  as an  $N \times 1$  vector containing  $N$  successive samples of the output of the  $m$ -th antenna sampled at a rate above or equal to the Nyquist rate for the P(Y) code.

$$\mathbf{x}_m(n) = [x_m(n), x_m(n-1), \dots, x_m(n-N+1)]^T \quad (1)$$

The  $NM \times 1$  space-time snapshot,  $\tilde{\mathbf{x}}(n)$ , is formed from concatenating  $\mathbf{x}_m(n)$ ,  $m = 1, 2, \dots, M$ , as

$$\tilde{\mathbf{x}}(n) = [\mathbf{x}_1(n); \mathbf{x}_2(n); \dots; \mathbf{x}_M(n)] \quad (2)$$

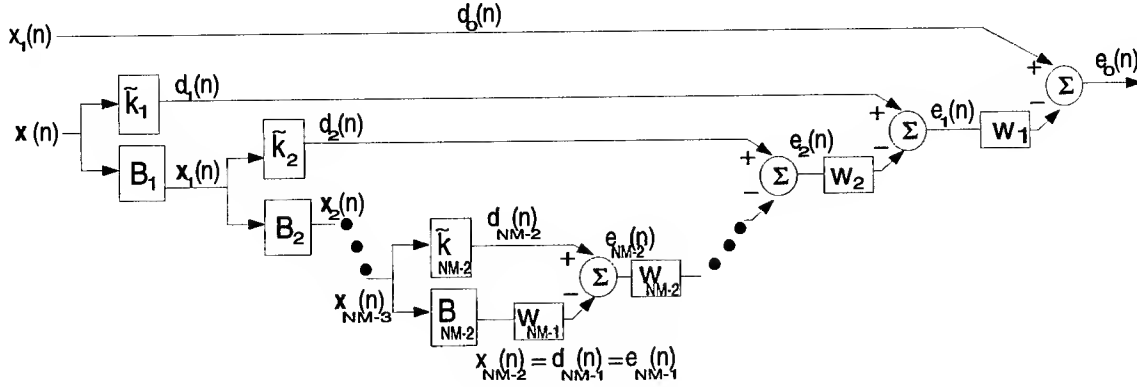


Figure 1. Nested chain of scalar Wiener filters for NM-1 joint space-time preprocessor.

where ; implies concatenating the vectors into a single column. Similarly, the  $N$  tap weights for the  $m$ -th antenna are placed as the components of an  $N \times 1$  vector as

$$\mathbf{h}_m = [h_m(0), h_m(1), \dots, h_m(N-1)]^T, \quad m = 1, 2, \dots, M \quad (3)$$

and the entire set of space-time weights is formed from a concatenation of  $\mathbf{h}_m$ ,  $m = 1, \dots, M$ , as

$$\mathbf{h} = [\mathbf{h}_1; \mathbf{h}_2; \dots; \mathbf{h}_M]. \quad (4)$$

The output power of the space-time preprocessor is

$$E\{|\mathbf{h}^H \tilde{\mathbf{x}}(n)|^2\} = \mathbf{h}^H \mathbf{K} \mathbf{h}, \quad \text{where: } \mathbf{K} = E\{\tilde{\mathbf{x}}(n) \tilde{\mathbf{x}}^H(n)\}. \quad (5)$$

Assume that the first antenna of the linear array is the reference antenna. To incorporate the unity weight constraint on the first tap of the reference antenna, define  $\mathbf{x}(n)$  as the  $(NM-1) \times 1$  sub-vector of  $\tilde{\mathbf{x}}(n)$  containing all but the first element of  $\tilde{\mathbf{x}}(n)$ . Similarly,  $\mathbf{h}_x$  is defined as the  $(NM-1) \times 1$  sub-vector of  $\mathbf{h}$  containing all but the first element of  $\mathbf{h}$ .

$$\tilde{\mathbf{x}}(n) = [x_1(n); \mathbf{x}(n)] \quad (6)$$

$$\mathbf{h} = [1; \mathbf{h}_x] \quad (7)$$

With these definitions, the power at the preprocessor output may be expressed as

$$E\{|\mathbf{h}^H \tilde{\mathbf{x}}(n)|^2\} = E\{|x_1(n) + \mathbf{h}_x^H \mathbf{x}(n)|^2\}. \quad (8)$$

Expressing the preprocessor output power in this fashion facilitates an adaptive filtering formulation where the output of the first tap of the reference antenna serves as the "desired" signal and the "error" signal is  $x_1(n) + \mathbf{h}_x^H \mathbf{x}(n)$ . As a result, LMS and/or RLS based adaptations are possible, as developed previously for the case of space-only processing [2].

### 3. DIMENSIONALITY REDUCTION VIA REDUCED-RANK METHODS

The disadvantage of space-time processing relative to space-only processing is the large dimensionality of the space-time correlation matrix relative to the spatial correlation

matrix. This translates into increased computational complexity and slower convergence. However, depending on the frequency and spatial distribution of the interferers, it may be possible to reduce the dimensionality. Reduction in dimensionality implies constraining the space-time weight vector to lie in a lower dimensional subspace. Defining an  $NM \times NM$  space-time correlation matrix  $\mathbf{K}$  (formed from  $M$  antennas with  $N$  taps per antenna), the original power minimization problem from [5] is

$$\begin{aligned} &\text{Minimize} \quad \mathbf{h}^H \mathbf{K} \mathbf{h} \\ &\text{subject to:} \quad \mathbf{h}^H \boldsymbol{\delta}_{NM} = 1 \end{aligned} \quad (9)$$

where  $\boldsymbol{\delta}_{NM}$  is the  $NM \times 1$  vector  $\boldsymbol{\delta}_{NM} = [0, 1, \dots, 0, \dots, 0]^T$  where the 1 is located in the  $NM$  position of the vector. We now seek to force the space-time weight vector to be in a particular reduced dimension subspace. That is let  $\mathbf{h} = \mathbf{T} \mathbf{h}_r$  where  $\mathbf{T}$  is the dimensionality reducing transformation matrix. Substitution of  $\mathbf{h} = \mathbf{T} \mathbf{h}_r$  into (9) allows one to rewrite the power minimization problem as

$$\begin{aligned} &\text{Minimize} \quad \mathbf{h}_r^H \mathbf{T}^H \mathbf{K} \mathbf{T} \mathbf{h}_r \\ &\text{subject to:} \quad \mathbf{h}_r^H \mathbf{T}^H \boldsymbol{\delta}_{NM} = 1 \end{aligned} \quad (10)$$

Using the method of Lagrange multipliers, the solution to (10) may be found by solving

$$\mathbf{T}^H \mathbf{K} \mathbf{T} \mathbf{h}_r = \alpha \mathbf{T}^H \boldsymbol{\delta}_{NM} \quad (11)$$

where  $\alpha$  is the Lagrange multiplier used to satisfy the unity weight constraint  $\mathbf{h}_r^H \mathbf{T}^H \boldsymbol{\delta}_{NM} = 1$ . It is easily shown that

$$\text{Minimum output power} = \frac{1}{\boldsymbol{\delta}_{NM}^H \mathbf{T} (\mathbf{T}^H \mathbf{K} \mathbf{T})^{-1} \mathbf{T}^H \boldsymbol{\delta}_{NM}}. \quad (12)$$

Since  $\mathbf{T}^H \mathbf{K} \mathbf{T}$  is Hermitian-symmetric, it follows that  $(\mathbf{T}^H \mathbf{K} \mathbf{T})^{-1}$  is Hermitian-symmetric, so that  $\alpha$  is real valued.

The reduced dimension transformation matrix  $\mathbf{T}$  can be found by techniques such as the cross-spectral metric (CS)[3] or principal-components (PC). A brief overview of these methods is necessary to motivate the use of the MSNWF. The space-time matrix  $\mathbf{K}$  can be spectrally decomposed

as  $\mathbf{K} = \sum_{i=1}^{NM} \lambda_i \mathbf{e}_i \mathbf{e}_i^H$ , where  $\lambda_i$  are the eigenvalues of  $\mathbf{K}$  indexed in descending order and  $\mathbf{e}_i$  are the corresponding eigenvectors. One can then seek dimensionality reduction through the transformation  $\mathbf{y}(n) = \mathbf{T}^H \mathbf{x}(n)$ , where

$\mathbf{T} = [\mathbf{e}_{i(1)} \mathbf{e}_{i(2)} \dots \mathbf{e}_{i(D)}]$  is an  $NM \times D$  matrix containing  $D < NM$  eigenvectors of  $\mathbf{K}$  and  $\{i(1), i(2), \dots, i(D)\}$  is a subset of the integers  $\{1, 2, \dots, NM\}$ . Given that the columns of  $\mathbf{T}$  are eigenvectors of  $\mathbf{K}$ , it follows that the  $D < NM$  eigenvectors of  $\mathbf{K}$  comprising  $\mathbf{T}$  can be selected as those which maximize the cross-spectral metric [3],[4] defined as

$$\delta_{NM}^H \mathbf{T} (\mathbf{T}^H \mathbf{K} \mathbf{T})^{-1} \mathbf{T}^H \delta_{NM} = \sum_{j=1}^D \frac{|\mathbf{e}_{i(j)}^H \delta_{NM}|^2}{\lambda_{i(j)}} \quad (13)$$

The principal-components technique would instead select the  $D$  largest eigenvectors of  $\mathbf{K}$  to form  $\mathbf{T}$ . Both techniques are quite computationally intensive since it is necessary to generate the eigenvectors of  $\mathbf{K}$  before finding the reduced-dimensioned matrix  $\mathbf{T}$  as well as compute  $(\mathbf{T}^H \mathbf{K} \mathbf{T})^{-1}$ . It was recently shown by [6] that the MSNWF generates a  $\mathbf{T}$  that may be expressed as

$$\mathbf{T} = [\delta_{NM}, \mathbf{K} \delta_{NM}, \dots, \mathbf{K}^{D-1} \delta_{NM}] \quad (14)$$

where again  $\mathbf{T}$  is an  $NM \times D$  matrix containing  $D < NM$  vectors associated with the  $D$  stages of the MSNWF. This formulation leads to a simple computation of  $\mathbf{T}$  as a function of the  $D$ -th stage chosen to truncate the MSNWF.

Once generating the particular  $\mathbf{T}$  associated with each reduced-rank method, it is possible to explore the effects of sample support associated with each  $\mathbf{T}$ . It was shown in [5] that the MSNWF outperformed both cross-spectral and principal-components in terms of jammer suppression as a function of rank. It is now of interest to examine the jammer suppression performance of each rank-reducing method as a function of sample support. This is illustrated in the simulations of Section 4. First, a brief development of the MSNWF algorithm is provided.

### 3.1. MSNWF Algorithm Development

Adaptive filtering schemes center upon a linear Minimum Mean Square Error (MMSE) estimation problem. In any linear MMSE problem, the optimum weight vector  $\mathbf{h}$  is the solution to the Wiener-Hopf equation

$$\mathbf{R}_{xx} \mathbf{h} = \mathbf{r}_{dx} \quad (15)$$

where  $\mathbf{R}_{xx}$  is the correlation matrix of the data and  $\mathbf{r}_{dx}$  is the cross-correlation vector between the data and the "desired" signal. The MSNWF represents a pioneering breakthrough in that it simultaneously achieves a convergence speed-up substantially better than that achieved with PC and a dramatically reduced computational burden relative to PC as well. Intuitively speaking, achieving the best of both worlds – faster convergence AND reduced computation – is made possible by making use of the information inherently contained in both  $\mathbf{R}_{xx}$  and  $\mathbf{r}_{dx}$  in choosing the reduced-dimension subspace that  $\mathbf{h}$  is constrained to lie within. In contrast, PC only makes use of the information embedded in  $\mathbf{R}_{xx}$ .

In our application here,  $\mathbf{R}_{xx} = \mathbf{K}$  and  $\mathbf{r}_{dx} = E\{x_1^*(n) \mathbf{x}(n)\}$ , assuming the unity weight constraint is at the first tap of the first antenna, for example. The MSNWF algorithm is summarized below[4]. As per the discussion at the end of Section 2, the "desired" signal  $d_0(n)$  is the output of the  $n$ -th tap at the  $m$ -th antenna.

- *Initialization:*  $d_0(n)$  and  $\mathbf{x}_0(n) = \mathbf{x}(n)$
- *Forward Recursion:* For  $k = 1, 2, \dots, D$ :

$$\begin{aligned} \mathbf{p}_k &= E\{d_{k-1}^*(n) \mathbf{x}_{k-1}(n)\} / \|E\{d_{k-1}^*(n) \mathbf{x}_{k-1}(n)\}\| \\ d_k(n) &= \mathbf{p}_k^H \mathbf{x}_{k-1}(n) \\ \mathbf{B} &= \mathbf{I} - \mathbf{p}_k \mathbf{p}_k^H \\ \mathbf{x}_k(n) &= \mathbf{B} \mathbf{x}_{k-1}(n) \end{aligned} \quad (16)$$

- *Backward Recursion:* For  $k = D, D-1, \dots, 1$ , with  $\epsilon_D(n) = d_D(n)$ :

$$\begin{aligned} w_k &= E\{d_{k-1}^*(n) \epsilon_k(n)\} / E\{|\epsilon_k(n)|^2\} \\ \epsilon_{k-1}(n) &= d_{k-1}(n) - w_k^* \epsilon_k(n) \end{aligned}$$

It follows that the matrix  $\mathbf{T} = [\mathbf{p}_1, \mathbf{p}_2, \dots, \mathbf{p}_D]$  contains orthonormal columns and that the reduced dimension  $D \times D$  correlation matrix  $\mathbf{T}^H \mathbf{K} \mathbf{T}$  is tri-diagonal [4]. The MSNWF is depicted in Figure 1 which clearly displays the multiple stages and nested structure. Operating in a  $D$ -dimensional space is tantamount to "cutting off" all stages below the  $D$ -th stage. The updating of the scalar weights  $w_k$  in Figure 1 may be effected through a simple LMS algorithm.

## 4. SIMULATIONS

Two scenarios are presented to illustrate the performance of the reduced-rank MSNWF in terms of nulling both wide-band and narrowband jammers while operating in a reduced-rank mode at low sample support. Consider  $M = N = 7$ . These definitions imply an  $M = 7$  element equi-spaced linear array with  $N = 7$  taps at each antenna. Although typical antenna arrays for GPS are two-dimensional (planar), circular, for example, or conformal, a linear array was used in this illustrative simulation example in order to have only one angular variable for display purposes. This allows the use of a single mesh or contour plot to display the two-dimensional Fourier Transform of the space-time weights obtained from a given run. In addition,  $N = 7$  is a very small number of taps to employ at each antenna. A significantly larger number is needed in practice in order to form sharp 'point-nulls' at the angle-frequency coordinates of a narrowband interferer and thereby minimize distortion to the GPS signal. The simulated preprocessor is constrained so that  $x_1(n)$  ( $1^{\text{st}}$  tap behind  $1^{\text{st}}$  antenna) is our reference signal, i.e.  $h_1(0) = 1$ . The other taps behind each antenna element form the column data vector  $\mathbf{x}(n)$  entering the stages of the MSNWF as illustrated in Figure 1. Table 1 summarizes the values used in the first scenario. Five of the six jammers for this simulation are narrowband jammers with different angles of arrival (AOAs). In both



scenarios, the narrowband jammers have different frequency offsets relative to the L1 frequency. Since we are assuming a 20 MHz receiver bandwidth at each antenna, the noise floor was determined to be approximately -128 dBW after filtering at each antenna. Recall the goal of power minimization is to drive the output power of the space-time beamformer as close to the noise floor as possible.

#### 4.1. Reduced Dimension Performance

Figures 2 and 3 plot the average power output of the space-time power minimization preprocessor based on the MSNWF as a function of subspace dimension or rank of the dimensionality reducing matrix transformation. The subspace dimension at which MSNWF approximately achieves the performance of the full-dimension ideal (asymptotic) Wiener filter is roughly the same in both scenarios, around eight. In contrast, Principal Components (PC) generally requires a subspace dimension equal to the number of degrees of freedom taken up by the jammers to achieve the same output power level. Each narrowband jammer takes up one degree of freedom. Each wideband jammer takes up  $N = 7$  degrees of freedom, where  $N$  is the number of taps per antenna. This is because the cancellation of a wideband jammer requires a spatial null, implying a null across the entire 20.46 MHz spectrum at its AOA. In Scenario 1, the jammers take up  $5 \times 1 + 1 \times 7 = 12$  degrees of freedom; in Scenario 2, the jammers take up  $1 \times 1 + 5 \times 7 = 36$  degrees of freedom.

#### 4.2. Low Sample Support Performance

Figures 4 and 5 examine the space-time snapshot sample support necessary to effectively null the jammers for each of the two scenarios simulated. The power output for each sample support level was averaged over 250 Monte Carlo trial runs. Each reduced-rank method used its respective ideal reduced dimension subspace matrix  $T$  in calculating the power output at each snapshot. Once the number of snapshots was equal to the rank for each reduced-rank method, the power output was calculated. The greatest differential in performance between the MSNWF and PC based methods is observed in Figure 5 corresponding to Scenario 2. In this case, Figure 3 and the above calculation dictate that PC needs to adapt in a 36-dimensional subspace, while the MSNWF need only adapt in a 10-dimensional space. This allows MSNWF to converge more rapidly than PC and CS. Note that the MSNWF is able to null the jammers effectively at low ranks with the added advantage of not requiring the computation of eigenvectors.

#### 4.3. Nulling Performance/Distortion Issues

Figures 6 and 7 display contour plots of the magnitude of the multi-dimensional Fourier Transform of the space-time weights obtained from the MSNWF with 40 space-time snapshots. For Scenario 1, Figure 6 displays a well-defined "point-null" at the angle-frequency coordinate of each narrowband jammer and a well-defined "line-null" along the arrival angle of the wideband jammer. For Scenario 2, Figure 7 displays a well-defined "point-null" at the angle-frequency coordinate of the one narrowband jammer and a well-defined "line-null" along the respective arrival angle

of each of the five the wideband jammers. As important, in both cases the response of the space-time beamformer is observed to be relatively flat away from the null locations.

### 5. CONCLUSION

The MSNWF preprocessor was shown to exhibit exceptional nulling performance for both wideband and narrowband jammers at low sample support and low rank. The reduced dimension subspace selected by the MSNWF exhibits rapid convergence in rank and sample support implying adaptive null tracking in a dynamic jamming environment. The MSNWF preprocessor was shown to outperform both principal-components and cross-spectral metric while operating at a lower rank and sample support.

Table 1: Simulation Parameters

Jammer Type	SNR	AOA Scen.1	AOA Scen.2	Bandwidth
Wideband	-100 dBW	20°	20°	20 MHz
Wideband	-110 dBW	—	0°	20 MHz
Wideband	-100 dBW	—	-20°	20 MHz
Wideband	-100 dBW	—	-40°	20 MHz
Wideband	-110 dBW	—	-60°	20 MHz
Jammer Type	SNR	AOA	AOA	Frequency
Narrowband	-100 dBW	60°	60°	-10 MHz
Narrowband	-100 dBW	15°	—	-5 MHz
Narrowband	-100 dBW	-10°	—	0 MHz
Narrowband	-100 dBW	-30°	—	5 MHz
Narrowband	-110 dBW	-55°	—	10 MHz

### REFERENCES

- [1] M. D. Zoltowski and A. S. Gecan, "Advanced Adaptive Null Steering Concepts for GPS," *Milcom '95*, vol. 3, pp. 1214-1218, 5-8 Nov. 1995.
- [2] A. S. Gecan and M. D. Zoltowski, "Power Minimization Techniques for GPS Null Steering Antennas," *Institute of Navigation (ION) Conference*, Palm Springs, CA, 13-15 Sept. 1995.
- [3] J. Scott Goldstein, I. S. Reed, and R. N. Smith, "Low-Complexity Subspace Selection for Partial Adaptivity," *Milcom '96*, vol. 2, pp. 597-601, 21-24 Oct. 1996.
- [4] J. Scott Goldstein, I. S. Reed, and L. L. Scharf, "A Multistage Representation of the Wiener Filter Based on Orthogonal Projections," *IEEE Trans. on Information Theory*, vol. 44, pp. 2943-2959, Nov. 1998.
- [5] W. L. Myrick, M. D. Zoltowski and J. Scott Goldstein, "Anti-Jam Space-Time Preprocessor for GPS Based on Multistage Nested Wiener Filter," *Milcom '99*.
- [6] M. L. Honig and W. Xiao, "Large System Analysis of Reduced-Rank Linear Interference Suppression," *Allerton '99*.

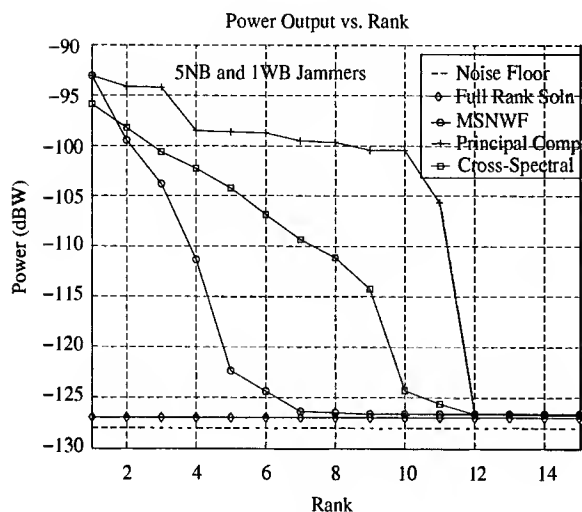


Figure 2. Power Output versus Rank (Scenario 1)  
Average Power Output of Preprocessor

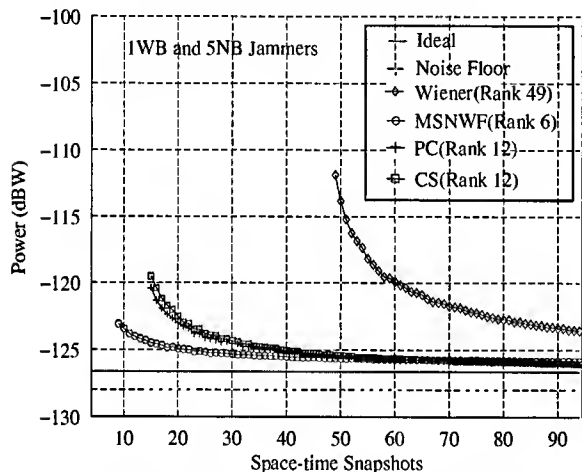


Figure 4. Power Output versus Snapshots (Scenario 1)  
Contour Plot of 2D DFT of Space-Time Weights

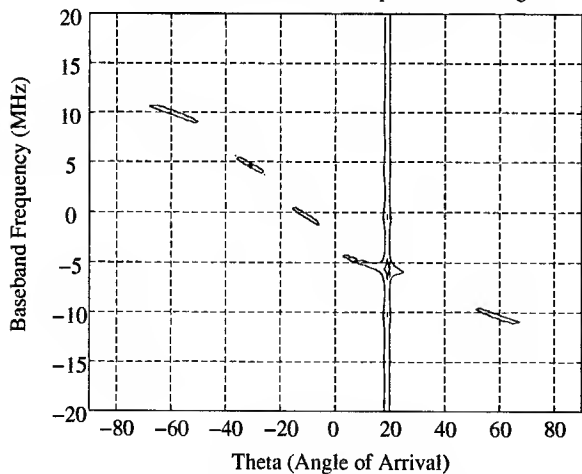


Figure 6. Contour Plot of 2D DFT (Scenario 1)

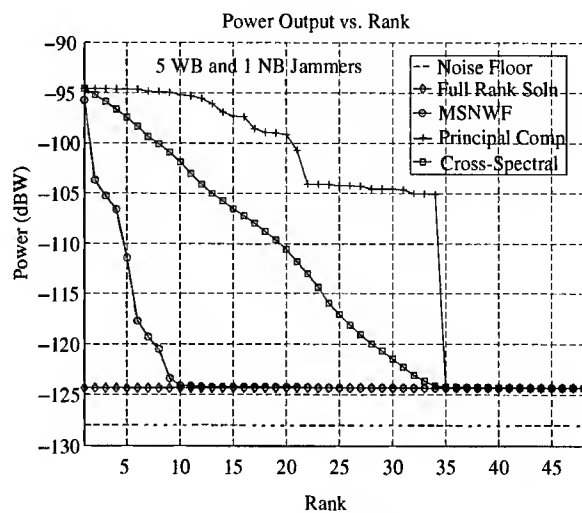


Figure 3. Power Output versus Rank (Scenario 2)  
Average Power Output of Preprocessor

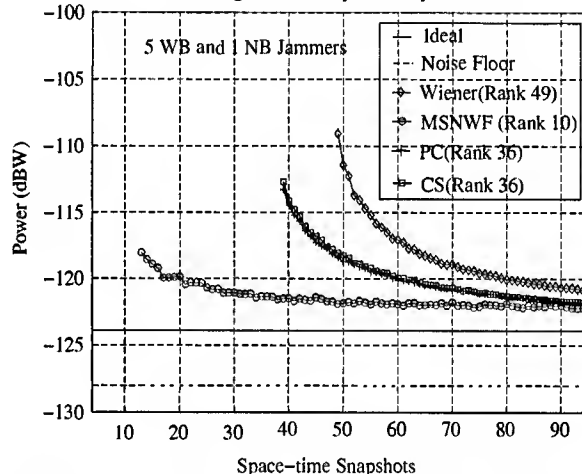


Figure 5. Power Output versus Snapshots (Scenario 2)  
Contour Plot of 2D DFT of Space-Time Weights

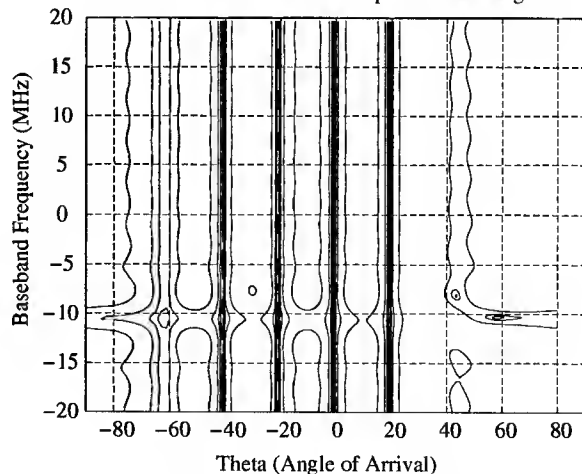


Figure 7. Contour Plot of 2D DFT (Scenario 2)

# JAMMER EXCISION IN SPREAD SPECTRUM USING DISCRETE EVOLUTIONARY-HOUGH TRANSFORM AND SINGULAR VALUE DECOMPOSITION

*Raungrong Suleesathira and Luis F. Chaparro*

Department of Electrical Engineering  
University of Pittsburgh, Pittsburgh, PA 15261, USA

## ABSTRACT

Jammer excision enhances interference immunity in direct sequence spread spectrum (DSSS) communications. In this paper, we propose an excision procedure, based on the discrete evolutionary and Hough transforms, for jammers composed of arbitrary chirps. The proposed instantaneous frequency (IF) estimation is done locally, without parameters, and it is recursively corrected. A singular value decomposition (SVD) of the dechirped signal allows us to synthesize the jammer locally, and then subtract it from the received signal. Localized processing, linearization of the IF estimate, recursive correction, and no problems due to cross-terms in the time-frequency distribution or in matching IF models make the proposed procedure efficient and practical. Also, SVD provides an efficient way to synthesize the jamming signal. The local IF estimation and the performance of the proposed exciser in DSSS systems are illustrated.

## 1. INTRODUCTION

Direct sequence spread spectrum (DSSS) techniques provide secure communication in an increasingly crowded spectrum. The advantages of DSSS are achieved by spreading the message so as to occupy a bandwidth in excess of the minimum needed. Despreading at the receiver with a synchronized replica of the spreading function permits recovery of the original message and reduces interferences. The received signal

$$r_k(n) = d_k p(n) + j_k(n) + \eta(n)$$

is composed of the product of the data bits  $d_k \in \{-1, 1\}$  and a pseudo white noise  $p(n) \in \{-1, 1\}$ ,  $0 \leq n \leq (N-1)$ , a jammer  $j_k(n)$  and a noise  $\eta(n)$  signals added during transmission. Although despreading  $r_k(n)$  recovers the original message while it spreads the jammer and noise, the performance of a DSSS system can fail if the power of the interferences is very strong. Excising

the jammers before despreading enhances the interference immunity.

Different methods have been proposed to mitigate broad-band jammers. The Wigner-Hough (WH) transform method [4] characterizes the jammer by a parametric model of its IF. Cross-terms and mismatching of the IF model hamper the method. Time-varying filtering and masking methods based on bilinear time-frequency (TF) distributions [2, 7] can excise jammers characterized by their instantaneous frequency, bandwidth and their support in the TF plane. A multiresolution method that uses a chirplet representation [3], and a procedure based on fractional Fourier transform [1] have also been proposed.

The authors recently presented an approach [9] for jammer excision based on the discrete evolutionary transform (DET) [11] and the Hough transform. In this paper, we exploit the advantages of local IF estimation, and by means of SVD obtain the significant singular values containing the jammer information. The local estimation permits us to consider multiple chirps in the jammer, and the SVD gives excellent synthesis of the jammer. Subtracting the synthesized jammer from the received signal before despreading enhances the interference immunity considerably.

## 2. DISCRETE EVOLUTIONARY-HOUGH TRANSFORM (DEHT)

### 2.1. Malvar-based DET

A non-stationary signal  $x(n)$  can be expressed as a sum of overlapping segments  $x_i(n)$ ,

$$x(n) = \sum_{i=0}^{I-1} x_i(n) \quad 0 \leq n \leq (N-1) \quad (1)$$

where  $I$  is the number of segments in which the signal is separated by Malvar windows  $\{v_i(n)\}$  having symmetrical overlaps at the partition points and such that

$\sum_i v_i^2(n) = 1$  (see Fig.1). The DET relates  $x_i(n)$  and its evolutionary kernel  $X_i(n, \omega_\ell)$ , [10, 11], as:

$$x_i(n) = \sum_{\ell=0}^{L_i-1} X_i(n, \omega_\ell) e^{j\omega_\ell n} \quad (2)$$

$$X_i(n, \omega_\ell) = \sum_{m=0}^{N-1} x(m) W_m^i(n, \ell) e^{-j\omega_\ell m} \quad (3)$$

where  $\omega_\ell = \frac{\pi}{L_i} \ell$ . The  $W_m^i(n, \ell)$  window is expressed in terms of the orthogonal Malvar wavelets  $\{u_{i\ell}(n) = v_i(n) \tilde{f}_{i\ell}(n)\}$  as

$$W_m^i(n, \ell) = u_{i\ell}(m) u_{i\ell}(n) e^{j\omega_\ell(m-n)} \quad (4)$$

The functions  $\{\tilde{f}_{i\ell}(n)\}$  are extensions of the orthogonal functions  $\{f_{i\ell}(n)\}$  given by

$$f_{i\ell}(n) = \sqrt{\frac{2}{L_i}} \cos\left(\frac{\pi}{L_i}(\ell + 0.5)(n - a_i)\right) \quad 0 \leq \ell \leq (L_i-1)$$

The Malvar expansion of  $x_i(n)$ , [10], is given by

$$x_i(n) = \sum_{\ell=0}^{L_i-1} c_{i\ell} u_{i\ell}(n) \quad (5)$$

where the coefficients  $c_{i\ell}$  are obtained by means of the orthogonality of the Malvar wavelets. The partition lengths  $\{L_i\}$  can be chosen independently. The criterion proposed in [5] find them by an entropy minimization. To consider different types of chirps, including sinusoids, we will extend their criterion.

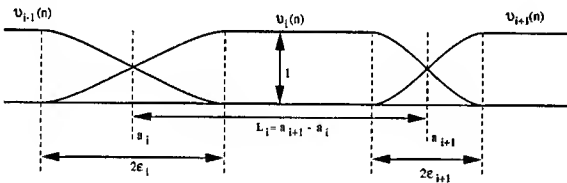


Figure 1: Typical Malvar window function

## 2.2. Optimal Windowing

To improve the entropy-based selection of the window lengths, we will use a zero-crossing rate and the rank of a Hankel matrix generated from the signal. These additional measures are especially useful when sinusoids are used as jammers.

As proposed in [5], minimizing the information cost

$$\lambda_i = - \sum_{\ell=0}^{L_i-1} |c_{i\ell}|^2 \log |c_{i\ell}|^2$$

is equivalent to minimizing the entropy of the  $\{c_{i\ell}\}$ . In some situations, the obtained partition can be improved by using the frequency content of the segment being considered. Such information can be obtained by a zero-crossing rate:

$$Z_i = \frac{1}{2L_i} \sum_{n=1}^{N-1} |\text{sgn}[x_{w_i}(n)] - \text{sgn}[x_{w_i}(n-1)]|$$

where  $x_{w_i}(n) = x(n)w_i(n)$ , and  $w_i(n)$  is a rectangular window of length  $L_i$  on  $(a_i, a_{i+1}]$  and  $\text{sgn}[\cdot]$  is the sign function. The rank of a Hankel matrix, [6], of  $x_{w_i}(n)$  denoted as

$$\mathbf{H}_i = \begin{bmatrix} x_{w_i}(0) & x_{w_i}(1) & \cdots & x_{w_i}(c-1) \\ x_{w_i}(1) & x_{w_i}(2) & \cdots & x_{w_i}(c) \\ \vdots & \vdots & \ddots & \vdots \\ x_{w_i}(L_i-c) & x_{w_i}(L_i-c+1) & \cdots & x_{w_i}(L_i-1) \end{bmatrix}$$

with  $(L_i - c + 1)$  rows and  $c$  columns, provides an estimate of the number of sinusoids in the segment. The value  $c$  is chosen greater than  $2r_i$ , where  $r_i$  is the number of sinusoidal components in the signal windowed by  $w_i(n)$ . After computing the above three measures in each partition, we use them to merge or split windows to obtain the best possible partition of  $x(n)$ .

## 2.3. Discrete Evolutionary-Hough Transform

Once the optimal lengths are chosen, upsampling is performed to achieve uniform frequency resolution in all segments. This is achieved by letting  $K$  be the least common multiple of the lengths, i.e.  $K = \text{LCM}\{L_i\}_{i \in [0, I-1]}$ . With identical frequency resolutions for every overlapped interval, the discrete evolutionary kernel  $X(n, \omega_k)$ , and the spectrum  $S(n, \omega_k)$  of the signal  $x(n)$  are given by

$$X(n, \omega_k) = \sum_{i=0}^{I-1} X_i(n, \omega_k)$$

$$S(n, \omega_k) = |X(n, \omega_k)|^2$$

The discrete evolutionary-Hough transform (DEHT) of  $x_{w_i}(n)$ , [9], is given by

$$\text{DEHT}(\theta, \rho, i) = \sum_{n,k} S(n, \omega_k) w_i(n) \delta(\rho - n \cos \theta - \omega_k \sin \theta)$$

where  $\delta(\cdot)$  is the Dirac delta function. The DEHT provides a linear estimate - characterized by the parameters  $(\rho, \theta)$  - of the IF of each of the chirps in a segment. The direct distance parameter  $\rho$  is the distance between the line appearing in  $S(n, \omega_k)w_i(n)$  and the origin. The inclination parameter  $\theta$  is the angle between  $\rho$  and the  $n$ -axis.

### 3. INSTANTANEOUS FREQUENCY ESTIMATION VIA DEHT

A problem in IF estimation using the Wigner-Hough transform is the mismatching between the actual IF and the used parametric models used for it. The linear IF estimator developed in [9] is local, recursive, and non-parametric, and valid for mono and multi-component signals.

Applying the Hough transform to the Malvar-based evolutionary spectrum of  $x_i(n)$  results in a piecewise linear characterization of the IF. This is due to the approximate sinusoidal representation obtained locally. It is also possible to recursively correct the estimate by processing the signal. Let  $\hat{\phi}_{iq}(n)$  be the initial instantaneous phase estimated for the  $q^{th}$ -component of

$$x_{w_i}(n) = \sum_p A_{ip} e^{j\phi_{ip}(n)}.$$

To improve the estimate  $\hat{\phi}_{iq}(n)$ , we first dechirp  $x_{w_i}(n)$  with it to get,

$$\begin{aligned} y_i^q(n) &= x_{w_i}(n) e^{-j\hat{\phi}_{iq}(n)} \\ &= A_{iq} e^{j[\phi_{iq}(n) - \hat{\phi}_{iq}(n)]} + \sum_{p \neq q} A_{ip} e^{j[\phi_{ip}(n) - \hat{\phi}_{iq}(n)]} \end{aligned}$$

Low-pass filtering  $y_i^q(n)$  with a narrow-band filter gives

$$\begin{aligned} z_i^q(n) &= A_{iq} e^{j[\phi_{iq}(n) - \hat{\phi}_{iq}(n)]} + e(n) \\ &= \tilde{A}_{iq} e^{j\tilde{\phi}_{iq}(n)} \end{aligned}$$

where  $e(n)$  is a small output with frequencies outside the filter bandwidth. Letting  $\tilde{\phi}_{iq}(n)$  be the phase of the filter output  $z_i^q(n)$ , we then obtain an improved estimate as

$$\hat{\phi}_{iq\text{new}}(n) = \hat{\phi}_{iq}(n) + \tilde{\phi}_{iq}(n)$$

We can then use this new estimate to dechirp  $x_{w_i}(n)$  again and find an improved estimate, repeating the process until the difference between the old and the new estimates is insignificant. The final estimate is then linearly fitted. The procedure is done locally and recursively for each of the signal components.

The performance of the local IF estimator is illustrated by considering first a sinusoidal FM signal. Its spectrum and the IF estimate are shown in Fig. 2(a)-(b); the final IF estimate (solid line) is very close to the actual one (dotted line). The dash-dotted line in the Fig. 2(b) is the piecewise linear IF obtained from the DEHT. As a second example, consider a signal consisting of a linear and a sinusoidal FM signals, with different constant amplitudes, embedded in noise (SNR 2.64 dB). The resulting estimates (solid line) are considerable improved over the initial estimates (dash-dotted line) as seen in Fig. 2(c)-(d).

### 4. APPLICATION TO JAMMER EXCISION IN DSSS COMMUNICATIONS

As suggested in [7], if the jammer is synthesized and subtracted from the received signal the performance of the despreading in DSSS is enhanced considerably. Assuming the jammer is composed of an arbitrary number of chirps with smooth IFs and time-varying amplitudes changing slowly in time, the proposed exciser is the one displayed in Fig. 3. After an estimate of the IF of one of the jammer components is obtained, this jammer component can be approximately synthesized by either low-pass filtering, or SVD of a modified Hankel matrix in the  $i^{th}$  segment. The modified Hankel matrix  $\mathbf{M}_i^q$ , of dimension  $L_i/2 \times (L_i + 2)/2$ , can be expressed in term of  $\mathbf{H}_i$  (generated from  $r_{ik}(n) = r_k(n)w_i(n)$ ) as

$$\mathbf{M}_i^q = \mathbf{H}_i e^{-j\hat{\phi}_{iq}(n)} \quad (6)$$

The rank of the modified Hankel matrix determines the number of chirps with the instantaneous phase  $\phi_{iq}(n)$  present in the signal. The most significant singular values will permit us to obtain a good synthesis of the jammer component after it is chirped using the estimated IF.

The number of significant singular values chosen is crucial in the jammer synthesis. To decide the number of singular values we use as a criterion a percentage of the energy of the dechirped received signal. Let  $E_{ik}^q$  be the energy of  $r_{ik}(n)e^{-j\hat{\phi}_{iq}(n)}$ ,  $\{\sigma_m\}$  be the singular values of  $\mathbf{M}_i^q$  and  $\{e_m\}$  be the eigenvalues of  $(\mathbf{M}_i^q)^* \mathbf{M}_i^q$  (the symbol  $*$  stands for the conjugate transpose). The matrix  $(\mathbf{M}_i^q)^* \mathbf{M}_i^q$  is symmetric with diagonal entries  $d_m$ ,  $m = 0, \dots, L_i/2$ . Observing that the sum of the first and the last diagonal entries of this symmetric matrix equals to the energy of the dechirped signal we have

$$\begin{aligned} E_{ik}^q &= d_0 + d_{L_i/2} \\ &= \text{Trace}[(\mathbf{M}_i^q)^* \mathbf{M}_i^q] - \sum_{m=1}^{L_i/2-1} d_m \quad (7) \end{aligned}$$

Given that  $\text{Trace}[(\mathbf{M}_i^q)^* \mathbf{M}_i^q] = \sum_m e_m$  [8], and that  $\sigma_m^2 = e_m$ , then Eq. (7) becomes

$$\begin{aligned} E_{ik}^q &= \sum_{m=0}^{L_i/2} e_m - \sum_{m=1}^{L_i/2-1} d_m \\ &= \sum_{m=0}^{L_i/2} \sigma_m^2 - \sum_{m=1}^{L_i/2-1} d_m \quad (8) \end{aligned}$$

If we let  $\sum_{m=1}^{L_i/2-1} d_m = \beta E_{ik}^q$  for some constant  $\beta$ , then

Eq. (8) can be rewritten as

$$E_{ik}^q = \frac{1}{1+\beta} \sum_{m=0}^{L_i/2} \sigma_m^2 \quad (9)$$

Once  $\beta$  is computed, the effect of choosing a certain number of singular values can be measured in terms of their contribution to the energy of the dechirped received signal.

To assess the performance of our procedure, we simulated DSSS received signals using 127 chips/bit, for various SNRs and two fix JSRs (jammer to signal ratios). The jammer is composed of a linear and a sinusoidal FM signals with different constant amplitudes just as in Fig. 2. Figures 4-5 show the probabilities of bit error when the received signal is jammed with JSRs of 26 and 34 dB. We consider 4 cases: when the received signal is without jammers; when no excision is performed before despreading; when using a lowpass filter, and when using the SVD. The results are improved after excising using the lowpass filter and SVD. However, the SVD method performs better in the case of stronger jammers, as is shown in Fig. 5. Figure 6(a)-(b) compares the excised signals using a low-pass filter and the SVD method to the received signal without jammers. As shown in Fig. 6(b), the low-pass filter method displays a large ripple at the boundaries of the segments.

## 5. CONCLUSIONS

Excision of a multi-component chirp jammer in DSSS communications is achieved by subtracting a synthesized version of it from the received signal. The jammer synthesis can be done by using lowpass filtering or SVD having estimates for the IFs. The IF estimator uses the DEHT to obtain a piecewise linear estimate, which can then be recursively corrected. From the results, it is shown that SVD seems to consistently perform better than lowpass filtering in the case of stronger jammers. Applying this excision procedure to frequency hopping spread spectrum systems will be further investigated.

## 6. REFERENCES

- [1] Akay O. and Boudreaux-Bartels F. G., "Broadband interference excision in spread spectrum communication systems via fractional Fourier transform," *Proc. Asilomar Conf. on Sig., Sys. and Comp.*, pp. 832-7, Nov. 1998.
- [2] Amin M. G. and Mandapati R. G., "Nonstationary interference excision in spread spectrum communications using projection filtering methods," *Proc. Asilomar Conf. on Sig., Sys. and Comp.*, pp. 827-31, Nov. 1998.

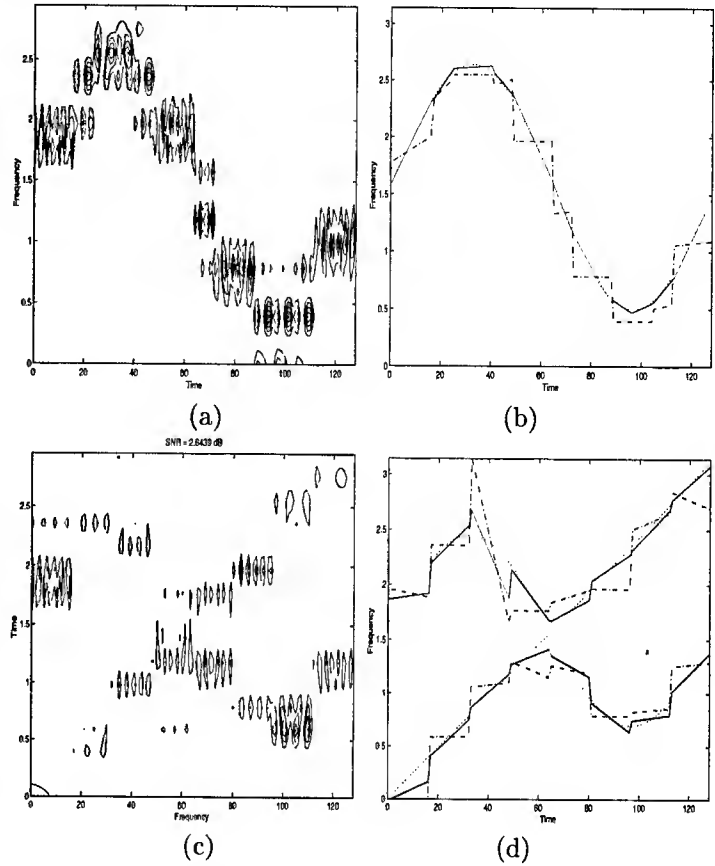


Figure 2: (a) and (c) Evolutionary spectra; (b) and (d) IF estimates; DEHT (dash-dotted), second iteration of correction (solid), actual IF (dashed)

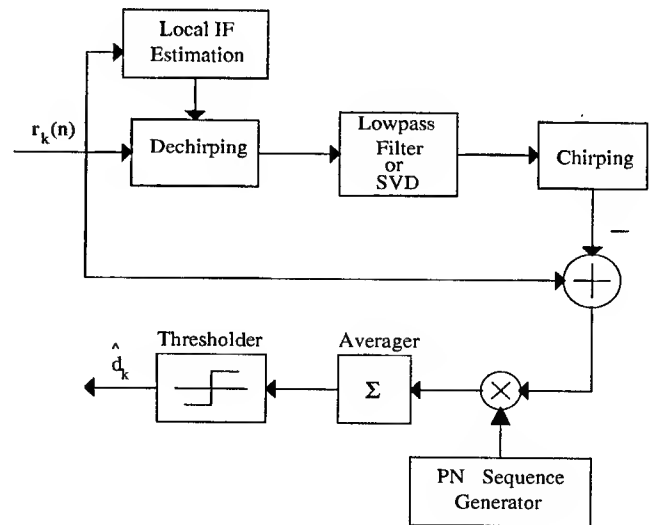


Figure 3: Diagram of the exciser using local instantaneous frequency estimation and lowpass filter or SVD

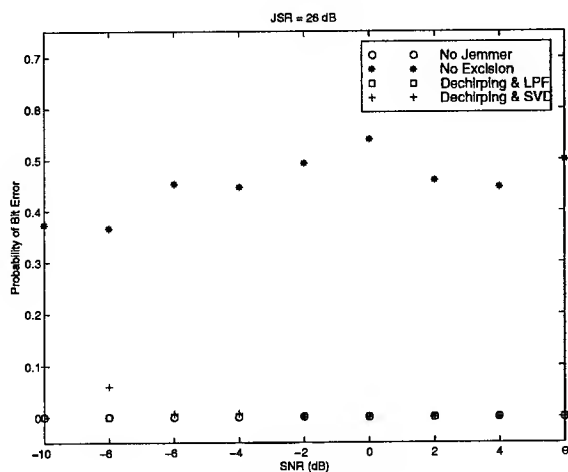


Figure 4: The probabilities of bit error versus SNRs under JSR of 26 dB.

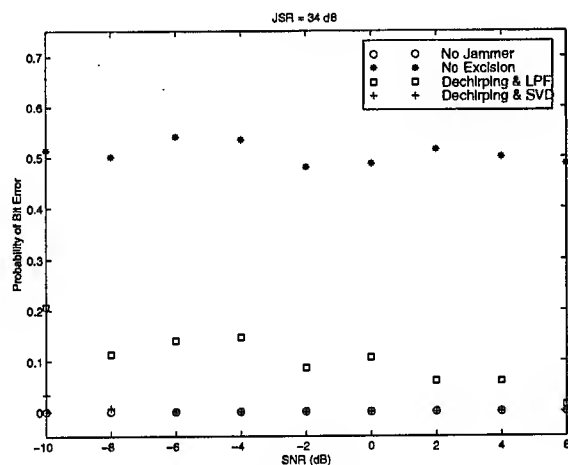
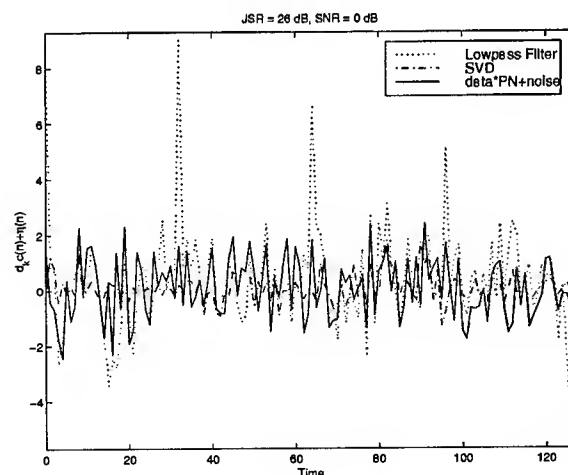
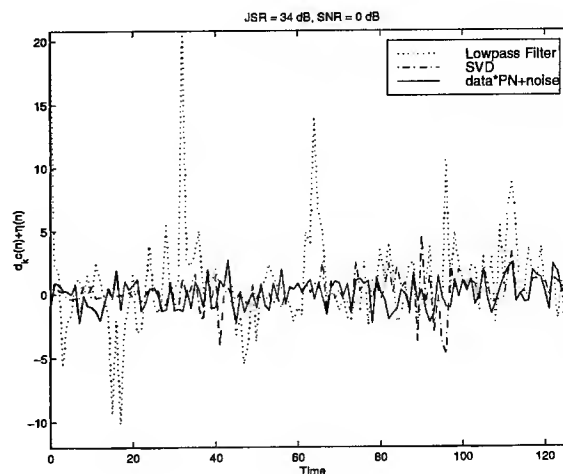


Figure 5: The probabilities of bit error versus SNRs under JSR of 34 dB



(a)



(b)

Figure 6: The comparison between exciser output and the  $d_k p(n) + \eta(n)$  by using the lowpass filtering and SVD at SNR = 0 dB and JSR are equal to (a) 26 and (b) 34 dB

- [3] Bultan A. and Akansu N. A., "A novel time-frequency exciser in spread spectrum communications for chirp-like interference," *Proc. ICASSP-98*, pp. 3265-68.
- [4] Barbarossa, S. and Scaglione A., "Adaptive time-varying cancellation of wideband interferences in spread spectrum communications based on time-frequency distributions," *IEEE Trans. Sig. Proc.*, pp. 957-65, Apr. 1999.
- [5] Coifman, R. R., and Wickerhauser, M. V., "Entropy-based algorithms for best basis selection," *IEEE Trans. on Info. Theory*, pp. 713-18, Mar. 1992.
- [6] DiMonte L.C., and Arun K.S., "Tracking the frequencies of superimposed time-varying harmonies," *IEEE Proc. ICASSP-90*, pp. 2539-42.
- [7] Lach R. S., Amin G. M. and Lindsey R. A., "Broad-band nonstationary interference excision for spread spectrum communications using time-frequency synthesis," *Proc. ICASSP-98*, pp. 3257-60.
- [8] Nobel B., *Applied Linear Algebra*, Prentice Hall, 1969.
- [9] Suleesathira, R. and Chaparro, L.F., "Interference mitigation in spread spectrum using discrete evolutionary and Hough transforms," *Accepted for presentation at ICASSP-00*.
- [10] Suleesathira, R., Chaparro, L. F. and Aydin, A., "Discrete evolutionary transform for time-frequency analysis," *Proc. Asilomar Conf. on Sig., Sys. and Comp.*, pp. 812-16, Nov. 1998.
- [11] Suleesathira, R., Chaparro, L. F. and Aydin, A., "Discrete evolutionary transform for time-frequency signal analysis," *Accepted for publication, Journal of the Franklin Institute, Special Issue in Time-Frequency*.

# SPATIAL AND TEMPORAL PROCESSING OF GPS SIGNALS

Ping Xiong\*, Michael J. Medley<sup>+</sup> and Stella N. Batalama\*

\*Department of Electrical Engineering  
State University of New York at Buffalo  
Buffalo, NY 14260

<sup>+</sup>Air Force Research Laboratory, IFGC  
525 Brooks Rd  
Rome, NY 13441

## ABSTRACT

In this paper we address the problem of navigation data demodulation by an adaptive GPS receiver that utilizes a bank of single-satellite linear-tap-delay filters and employs antenna-array reception. The presence of an antenna array allows the receiver to operate in the spatial domain in addition to the temporal (code) domain. We investigate disjoint-domain as well as joint-domain space-time GPS signal processing techniques and we consider design criteria of conventional matched-filter (MF) type, minimum-variance-distortionless-response (MVDR) type and auxiliary-vector (AV) type. The proposed structures utilize filters that operate at a fraction of the navigation data bit period (1 msec) and are followed by soft-decision detectors. Soft decisions taken over a navigation data bit period are then combined according to a simple combining rule. Simulation results illustrate the bit-error-rate (BER) performance of the investigated design alternatives.

## 1. INTRODUCTION

The Global Positioning System (GPS), originally developed for military use, has received a lot of attention recently for use in civilian applications such as aviation, agriculture, land-vehicle navigation, surveying and mapping, to name a few [1], [2]. The GPS system employs direct-sequence spread-spectrum (DS-SS) signaling. Each satellite is assigned a coarse acquisition (C/A) code and a precision (P) code. The C/A-code is a Gold sequence with chipping rate at 1.023 Mcips/sec and period 1 msec (or code-length 1023 chips), while the P-code is a pseudorandom code with chipping rate at 10.23 Mcips/sec and period one week. The

THIS WORK WAS SUPPORTED IN PART BY THE NATIONAL SCIENCE FOUNDATION UNDER GRANT CCR-9805359 AND IN PART BY THE DEFENSE ADVANCED RESEARCH PROJECTS AGENCY (DARPA) AND AIR FORCE RESEARCH LABORATORY, USAF, UNDER AGREEMENT NUMBER F30602-00-1-0520. THE U.S. GOVERNMENT IS AUTHORIZED TO REPRODUCE AND DISTRIBUTE REPRINTS FOR GOVERNMENTAL PURPOSES NOTWITHSTANDING ANY COPYRIGHT ANNOTATION THEREON. THE VIEWS AND CONCLUSIONS CONTAINED HEREIN ARE THOSE OF THE AUTHORS AND SHOULD NOT BE INTERPRETED AS NECESSARILY REPRESENTING THE OFFICIAL POLICIES OR ENDORSEMENTS, EITHER EXPRESSED OR IMPLIED, OF THE DEFENSE ADVANCED RESEARCH PROJECTS AGENCY (DARPA), THE AIR FORCE RESEARCH LABORATORY, OR THE U.S. GOVERNMENT.

C/A and P code are modulated by binary navigation data (at 50 bps) and then multiplexed in phase quadrature to form the satellite signal. In this paper we focus on the C/A component of the transmitted GPS signal (or as commonly stated, we assume perfect separation of the C/A and P-signal component at the receiver).

The working principle of the GPS system is very simple. The position of a GPS receiver is modeled as a four-dimensional vector (three coordinates correspond to the spatial position of the receiver and one is related to receiver timing). Estimation of the position can be achieved by utilizing the signals of a minimum of four satellites.

The signal captured by a GPS receiver is the aggregate of the GPS signals of the satellites that are currently in view, their multipaths, additive white Gaussian noise (AWGN), possible intelligent hostile spread spectrum (SS) interference (spoofing) and/or narrowband interference.

The component of the received signal that is due to the GPS signals of the satellites currently in view is the superposition of very low correlated SS signals. However an earth-based intelligent SS interferer/spoofers, who knows the satellite C/A-code can mimic the signal of interest and thus contribute highly correlated (with signal of interest) interference.

In this paper we address the problem of navigation data demodulation by an adaptive GPS receiver that utilizes a bank of single-satellite linear-tap-delay filters and employs antenna-array reception. The presence of an antenna-array allows the receiver to operate in the spatial domain in addition to the temporal (code) domain. We investigate disjoint-domain as well as joint-domain space-time GPS-signal processing techniques and we consider design criteria of the form of conventional matched-filter (MF) as well as interference suppressing minimum-variance-distortionless-response (MVDR) [3] and auxiliary-vector (AV) filtering [4]. The proposed structures utilize filters that operate at one twentieth of the navigation data bit period. Soft pre-detection measurements taken over a navigation data bit period are then combined according to a simple combining rule for further BER performance improvements.

## 2. SYSTEM MODEL

The C/A and P-signal component are assumed to be perfectly separated at the receiver. Then, the aggregate C/A component of the received signal can be viewed as an asynchronous DS-SS system with  $K$  SS signals in the presence of AWGN. Since the satellite navigation information bit



rate is 50 bps, it is equivalent to say that the C/A code of each GPS signal repeats itself 20 times and during this period it is modulated by the same data bit. In this context, the contribution of the  $k$ th transmitted GPS signal,  $k = 1, \dots, K$ , over  $T = LT_c$  secs is given by  $u_k(t) = \sum_i b_k(i) \sqrt{E_k} s_k(t - iT) e^{j2\pi f_c t}$ , where  $b_k(i) \in \{-1, +1\}$  is the  $i$ th transmitted data bit such that  $b_k(i) = b_k(i + m)$ ,  $m = 1, \dots, 19$ , and  $b_k(i)$  is independent of  $b_k(j)$  for all  $|i - j| > 20$ . Also,  $E_k$  and  $\phi_k$  denote the energy and the carrier phase, respectively, of the  $k$ th signal (with carrier frequency  $f_c$ ), while  $s_k$  is the normalized  $k$ th C/A code given by  $s_k(t) = \sum_{l=0}^{L-1} d_k(l) \psi(t - lT_c)$ , where  $d_k(l) \in \{\pm 1/\sqrt{L}\}$ ,  $l = 0, \dots, L - 1$ , are the signature bit values of the  $k$ th SS signal and  $\psi(t)$  is the chip waveform of duration  $T_c = T/L$ .

The aggregate signal received at the input of a narrow-band uniform linear array of  $M$  antenna elements is given by  $\mathbf{x}_c(t) = \sum_{k=1}^K u_k(t - \tau_k) \mathbf{a}_k + \mathbf{n}(t)$ , where  $\mathbf{n}(t)$  is complex AWGN and  $\mathbf{a}_k$  denotes the array response vector (spatial signature) of the  $k$ th SS signal with elements defined by  $a_k(m) \triangleq e^{j2\pi(m-1) \frac{\sin \theta_k d}{\lambda}}$ ,  $m = 1, \dots, M$  (in which  $\theta_k$  denotes the angle of arrival of the  $k$ th signal,  $\lambda$  is the carrier wavelength,  $d$  is the inter-element spacing-usually  $d = \lambda/2$ ).

After carrier demodulation, the received signal is given by

$$\mathbf{x}(t) = \sum_i \sum_{k=1}^K b_k(i) \sqrt{E_k} s_k(t - iT - \tau_k) e^{-j2\pi f_c \tau_k} \mathbf{a}_k + \mathbf{n}(t). \quad (1)$$

Without loss of generality, we assume chip synchronization at the reference antenna element ( $m = 1$ ) with the SS signal of interest, say *Signal 1*, and we also assume that  $0 \leq \tau_k < T$ ,  $k = 2, \dots, K$ . After conventional chip matched filtering and sampling at the chip rate  $1/T_c$ , we can visualize the space-time data samples associated with  $b_1(i)$  in the form of an  $M \times L$  matrix  $\mathbf{X}_{M \times L}(i) = [\mathbf{x}_{M \times 1}(iL) \quad \mathbf{x}_{M \times 1}(iL + 1) \cdots \mathbf{x}_{M \times 1}(iL + L - 1)]$  where the column vector  $\mathbf{x}_{M \times 1}(iL + j)$  is given by

$$\mathbf{x}_{M \times 1}(iL + j) = \sum_{k=1}^K b_k(i) \sqrt{E_k} d_k(j - \tau_k/T_c) e^{-j2\pi f_c \tau_k} \mathbf{a}_k + \mathbf{n}_{M \times 1}(iL + j), \quad j = 0, \dots, L - 1. \quad (2)$$

In the following, we pursue one-shot detection of the bit of interest  $b_1(i)$ , and we drop the index  $i$  for simplicity in notation.

In this work we focus on the detection of the navigation data of a single satellite. In this context GPS signals of other satellites currently in view as well as intelligent SS "GPS-looking" jamming signals are treated comprehensively as SS interference. The differences between the former and the latter lie in their corresponding signature cross-correlation level with the satellite signal of interest as well as their power level.

### 3. SPACE AND TIME PROCESSING ALTERNATIVES

In this section we investigate disjoint and joint domain filtering configurations for GPS signal processing. The disjoint configurations are formed by the cascade of a space filter followed by a time filter (S-T), or by a time filter followed by a space filter (T-S). In this context, disjoint

domain receiver design is a two stage process with the second stage being conditioned on the design of the first stage. The design optimization criterion imposed at either stage (regardless of the space or time nature of the corresponding filter) that is of interest in this work is of MF-type, MVDR-type or AV-type. The design optimization criterion for each stage can be selected independently and is usually dictated by considerations of simplicity in implementation, computational complexity and performance. Due to the lack of space we present only the studies for S-T configurations.

#### 3.1. Disjoint Space-Time (S-T) Configuration

For an arbitrary linear space processor,  $\mathbf{f}_s$ , and an arbitrary linear time processor,  $\mathbf{f}_t$ , the decision on the information bit of the signal of interest  $b_1$  is given by the following expression:

$$\hat{b}_1 = \text{sgn}(\text{Re}\{\mathbf{f}_t^H (\mathbf{f}_s^H \mathbf{X})^T\}) = \text{sgn}(\text{Re}\{\mathbf{f}_s^H \mathbf{X} \mathbf{f}_t^*\}) \quad (3)$$

where  $\text{sgn}(\cdot)$  identifies the sign operation and  $\text{Re}\{\cdot\}$  extracts the real part of a complex number. In the following we present the different forms that  $\mathbf{f}_s$  and  $\mathbf{f}_t$  may assume.

##### Spatial Matched-Filtering (sMF)

The spatial matched filter is the filter matched to the array response vector,  $\mathbf{a}_1$ , of the signal of interest, that is

$$\mathbf{f}_{sMF} = \mathbf{a}_1/M. \quad (4)$$

We observe that  $E_{b_1}\{\mathbf{X}(b_1 \mathbf{d}_1)\} = \sqrt{E_1} \mathbf{a}_1$ , where  $\mathbf{d}_1$  is the code of the signal of interest and the statistical expectation operation  $E\{\cdot\}$  is taken with respect to  $b_1$  only.  $\square$

##### Spatial MVDR Filtering (sMVDR)

The MVDR filter is designed to minimize the filter output variance subject to the constraint that the filter remains distortionless in the direction of the signal of interest  $\mathbf{a}_1$  [3]. The filter is given by

$$\mathbf{f}_{sMVDR} = \frac{\mathbf{R}_s^{-1} \mathbf{a}_1}{\mathbf{a}_1^H \mathbf{R}_s^{-1} \mathbf{a}_1} \quad (5)$$

where  $\mathbf{R}_s$  denotes the  $M \times M$  covariance matrix of the columns of  $\mathbf{X}_{M \times L}$  (i.e., the correlation of the spatial input data).  $\square$

##### Spatial Auxiliary-Vector Filtering (sAV)

For a given spatial covariance matrix  $\mathbf{R}_s$ , the theory of auxiliary vector (AV) filtering [4], [5] can be applied to the spatial domain to provide a sequence of spatial linear filters that are distortionless in the vector direction of interest  $\mathbf{a}_1$  and can be obtained by the following recursion:

$$\mathbf{f}_{sAV}(0) = \frac{\mathbf{a}_1}{\|\mathbf{a}_1\|^2} \quad (6)$$

For  $n = 1, 2, \dots$

$$\mathbf{g}_s(n) = \mathbf{R}_s \mathbf{f}_{sAV}(n-1) - \frac{\mathbf{a}_1^H \mathbf{R}_s \mathbf{f}_{sAV}(n-1) \mathbf{a}_1}{\|\mathbf{a}_1\|^2} \quad (7)$$

$$\mu_s(n) = \frac{\mathbf{g}_s^H(n) \mathbf{R}_s \mathbf{f}_{sAV}(n-1)}{\mathbf{g}_s^H(n) \mathbf{R}_s \mathbf{g}_s(n)} \quad (8)$$

$$\mathbf{f}_{sAV}(n) = \mathbf{f}_{sAV}(0) - \sum_{i=1}^n \mu_s(i) \mathbf{g}_s(i) \quad (9)$$

As shown in [5], given infinite data, the sequence of auxiliary vector filters converges to  $\mathbf{f}_{sMVDR}$  ( $\mathbf{f}_{sAV}(n) \rightarrow \mathbf{f}_{sMVDR}$  as  $n \rightarrow \infty$ ).

□

Finally, the time processor of the S-T configuration is a linear filter of dimension  $L \times 1$  that takes as input the space-processor output. Similar to the optimization criteria used in the design of the space processor, the temporal filter can be of MF-type, MVDR-type or AV-type.

#### Temporal Matched-Filtering (tMF)

The temporal MF is given by

$$\mathbf{f}_{tMF} = \mathbf{d}_1 \quad (10)$$

or equivalently,  $\mathbf{f}_{tMF} = E_{b_1}\{\mathbf{y}b_1\}$ , where  $\mathbf{y}$  is the output of the spatial processor.

□

#### Temporal MVDR-Filtering (tMVDR)

The temporal MVDR filter can be shown to be equal to

$$\mathbf{f}_{tMVDR} = \frac{\mathbf{R}_t^{-1} \mathbf{d}_1}{\mathbf{d}_1^H \mathbf{R}_t^{-1} \mathbf{d}_1} \quad (11)$$

where  $\mathbf{R}_t$  is the  $L \times L$  covariance matrix of the spatial filter output  $\mathbf{y}$  and thus it depends on the type of first-stage spatial processing.

□

#### Temporal Auxiliary-Vector Filtering (tAV)

The sequence of auxiliary vector filters in the time domain can be obtained as follows,

$$\mathbf{f}_{tAV}(0) = \mathbf{d}_1 \quad (12)$$

For  $n=1, 2, \dots$

$$\mathbf{g}_t(n) = \mathbf{R}_t \mathbf{f}_{tAV}(n-1) - \mathbf{d}_1^H \mathbf{R}_t \mathbf{f}_{tAV}(n-1) \mathbf{d}_1 \quad (13)$$

$$\mu_t(n) = \frac{\mathbf{g}_t^H(n) \mathbf{R}_t \mathbf{f}_{tAV}(n-1)}{\mathbf{g}_t^H(n) \mathbf{R}_t \mathbf{g}_t(n)} \quad (14)$$

$$\mathbf{f}_{tAV}(n) = \mathbf{f}_{tAV}(0) - \sum_{i=1}^n \mu_t(i) \mathbf{g}_t(i) \quad (15)$$

and  $\mathbf{R}_t$  denotes the covariance matrix of the spatial filter output  $\mathbf{y}$ . The following theorem provides a performance comparison in terms of output signal-to-interference-plus-noise-ratio (SINR) of the configurations presented above for two SS signal case. The proof is omitted due to the lack of space.

**Theorem 1** Let  $\mathbf{a}_i$  and  $\mathbf{d}_i$  be the spatial and temporal signature of user- $i$ ,  $i = 1, 2$ , of length  $M$  and  $L$ , respectively. Define  $\rho \triangleq \mathbf{d}_1^H \mathbf{d}_2$ ,  $\eta \triangleq |\frac{\mathbf{a}_1^H \mathbf{a}_2}{M}|$ . Let also  $E_i$  denote the signal-to-noise-ratio (SNR) of user- $i$ , and  $\text{SINR}_{(sXX/tYY)}$  or  $\text{SINR}_{(tYY/sXX)}$  denote the output SINR of an S-T or T-S configuration, that utilizes an  $XX$ -type space filter and a  $YY$ -type time filter. Then

#### A. Space-Time configuration

- (i)  $\text{SINR}_{(sMF/tMF)} < \text{SINR}_{(sMF/tMVDR)}$ , for any  $\rho, \eta$
- (ii)  $\text{SINR}_{(sMVDR/tMF)} < \text{SINR}_{(sMVDR/tMVDR)}$ , for any  $\rho, \eta$
- (iii) A loose sufficient condition for  $\text{SINR}_{(sMVDR/tMVDR)}$

$< \text{SINR}_{(sMF/tMVDR)}$  is

$$1 + ME_2\eta^2(1 - \rho^2) \geq \left[1 + \frac{ME_2\eta^2(1 - \rho^2)}{1 + \frac{ME_2}{L}(2 + \frac{ME_2}{L})(1 - \eta^2)}\right]^2 \left[1 + \frac{ME_2}{L}(1 - \eta^2)\right]^2. \quad (16)$$

#### B. Time-Space configuration

- (i)  $\text{SINR}_{(tMF/sMF)} < \text{SINR}_{(tMF/sMVDR)}$ , for any  $\rho, \eta$
- (ii)  $\text{SINR}_{(tMVDR/sMF)} < \text{SINR}_{(tMVDR/sMVDR)}$ , for any  $\rho, \eta$
- (iii) A loose sufficient condition for  $\text{SINR}_{(tMVDR/sMVDR)}$

$< \text{SINR}_{(tMF/sMVDR)}$  is

$$1 + ME_2\rho^2(1 - \eta^2) \geq \left[1 + \frac{ME_2\rho^2(1 - \eta^2)}{1 + E_2(2 + E_2)(1 - \rho^2)}\right]^2 \left[1 + E_2(1 - \rho^2)\right]^2. \quad (17)$$

#### C. Space-Time versus Time-Space configuration

- (i)  $\text{SINR}_{(sMF/tMF)} = \text{SINR}_{(tMF/sMF)}$ , for any  $\rho, \eta$
- (ii)  $\text{SINR}_{(sMVDR/tMF)} < \text{SINR}_{(tMF/sMVDR)}$ , for any  $\rho, \eta$
- (iii)  $\text{SINR}_{(tMVDR/sMF)} < \text{SINR}_{(sMF/tMVDR)}$ , for any  $\rho, \eta$

□

### 3.2. Joint Domain Filtering

In joint domain processing, to avoid cumbersome 2-D operations and notations, we vectorize the matrix  $\mathbf{X}_{M \times L}$  by stacking all columns in the form of a vector  $\mathcal{X}_{ML \times 1} = \text{Vec}\{\mathbf{X}_{M \times L}\}$ . In the following,  $\mathcal{X}$  denotes the joint space-time data in the  $C^{ML}$  complex vector space that constitutes the input to a joint space-time linear filter  $\mathbf{w}$  to be designed according to MF, MVDR or AV processing principles.

#### Joint Domain Matched Filtering (JMF)

The joint space-time matched filter  $\mathbf{w}_{JMF}$  for the signal of interest (Signal 1) is equal to the joint space-time signature of the signal of interest, i.e., the Kronecker product  $\mathbf{v}_1 = (\mathbf{d}_1 \otimes \mathbf{a}_1)/M$ , where  $\mathbf{d}_1$  and  $\mathbf{a}_1$  are the temporal signature and spatial signature (steering vector) of the signal of interest, respectively (JMF is equivalent to sMF/tMF and tMF/sMF configurations).

We note that JMF is optimum only when the channel interference plus noise is white Gaussian which is not the case for most practical SS communication systems. Indeed, non-orthogonal multiple access interferers as well as highly correlated (with the signal of interest) intentional jammers may render the JMF receiver obsolete. A remedy for the latter situation is to proceed with the design of interference suppressing receivers such as the MVDR receiver or the AV receiver that are presented below.

#### Joint Domain MVDR Filtering (JMVDR)

The joint domain MVDR filter is designed to minimize its output energy and simultaneously be distortionless toward the joint space-time signature of the signal of interest  $\mathbf{v}_1$ . It is given by the following expression:

$$\mathbf{w}_{JMVDR} = \frac{\mathbf{R}^{-1} \mathbf{v}_1}{\mathbf{v}_1^H \mathbf{R}^{-1} \mathbf{v}_1} \quad (18)$$

where  $\mathbf{R} = E\{\mathcal{X}\mathcal{X}^H\}$  is the covariance matrix of the space-time input data vector.

Joint domain auxiliary-vector filter design provides a sequence of joint domain auxiliary vector filters that are distortionless toward the joint space-time signature of the signal of interest  $\mathbf{v}_1$  and can be obtained by the following recursion:

$$\mathbf{w}_{JAV}(0) = \frac{\mathbf{v}_1}{\|\mathbf{v}_1\|^2} \quad (19)$$

For  $n=1, 2, \dots$

$$\mathbf{g}(n) = \mathbf{R}\mathbf{w}_{JAV}(n-1) - \frac{\mathbf{v}_1^H \mathbf{R}\mathbf{w}_{JAV}(n-1)\mathbf{v}_1}{\|\mathbf{v}_1\|^2} \quad (20)$$

$$\mu(n) = \frac{\mathbf{g}^H(n)\mathbf{R}\mathbf{w}_{JAV}(n-1)}{\mathbf{g}^H(n)\mathbf{R}\mathbf{g}(n)} \quad (21)$$

$$\mathbf{w}_{JAV}(n) = \mathbf{w}_{JAV}(0) - \sum_{i=1}^n \mu(i)\mathbf{g}(i) \quad (22)$$

### 3.3. GPS Filter Output Combining

To take advantage of the redundancy introduced by utilizing  $C/A$  codes of period equal to a fraction of the information bit period, and still maintain low-order filtering (equal to the  $C/A$  code length), combining methods such as *selective combining* (SC), *equal gain combining* (EGC) or *maximum ratio combining* (MRC) can be used to further improve the receiver BER performance. In this paper we utilize EGC because of its simplicity in implementation.

### 3.4. Filter Estimation Considerations

The developments so far involved filter alternatives of MF, MVDR and AV type. Besides the fixed MF-type structure, all others are adaptive in nature and have been presented/formulated under ideal conditions, that is, under the assumption that the space or time or space-time covariance matrix  $\mathbf{R}$  involved is known. In practice, however,  $\mathbf{R}$  is unknown and it is sample-average estimated by a data record of finite size. When  $\mathbf{R}$  is substituted by the sample-average estimated  $\hat{\mathbf{R}}$  then the ideal receiver expressions in (5), (9), (11), (15), (18) and (22) assume their estimated versions. In this context, the AV algorithm, produces a sequence of MVDR filter estimators of the form  $\hat{\mathbf{w}}(0), \hat{\mathbf{w}}(1), \dots$ . This sequence has been extensively studied in [5] and shown to offer the means for effective control over the filter estimator bias versus covariance trade-off. As a result, adaptive filter estimators from this class have been seen to easily outperform in mean-square estimation error the (constraint) LMS, sample-matrix-inversion (SMI) and RLS type adaptive filter implementations. These operational characteristics of the AV filter estimators place them favorably in terms of GPS receiver implementation when interference suppression with short data records is the objective. Simulation comparisons in the following section illustrate how the above observations translate into superior BER performance.

## 4. NUMERICAL AND SIMULATION COMPARISONS

We consider the GPS signal model in (1) for a system with  $M = 2$  antenna elements and spreading gain  $L = 1,023$ .

In all cases we assume the presence of 4 satellite signals with fixed  $C/A$  Gold codes, as well as the presence of one or two high power spread spectrum jammers (spoofers) that exhibit code cross-correlation with the signal of interest (*Signal 1*) approximately 0.1 and 0.2, respectively. The angles of arrival of the satellite and jamming signals are randomly generated according to a uniform distribution in  $(-\pi/2, \pi/2)$ .

The simulation/numerical studies in this section evaluate the BER performance of the GPS receiver as a function of either the SNR of the signal of interest or the data record size. All BERs are analytically evaluated and the results are averages over 100 independent space-time channels.

For the BER versus SNR studies, the signal of interest SNR varies from 0 dB (weak signal) to 15 dB (normal strength signal). The SNRs of the other satellite signals are fixed at 15 dB while the jammer's SNR is fixed at 30 dB and the AWGN variance is taken equal to 1. In Figure 1 the BER versus the SNR of *Signal 1* is shown for different disjoint and joint estimated receiver configurations in the presence of one high power SS jammer.

Figures 2-4 plot the BER versus the data record size in the presence of two jammers. Figure 2 plots the BER of the estimated disjoint S-T MF and MVDR type configurations while Figure 3 involves receiver configurations that utilize an auxiliary vector filter in the first stage or the second stage or both stages. Figure 4 plots the BER versus the data record size for the estimated joint-domain configurations. The SNR of *Signal 1* in Figures 2-4 is fixed at 15 dB.

Figures 1-4 illustrate the performance gains when non-MF-type signal processing is performed by the GPS receiver and do not consider combining. Additional performance gains obtained through EGC combining are illustrated in Figure 5 where we plot the BER as a function of data record size for the best receiver configuration of previous (Figs. 2-4) studies.

## REFERENCES

- [1] B. W. Parkinson, J. J. Spilker, P. Axelrad and P. Enge eds, *Global Positioning System: Theory and Applications I*, American Institute of Aeronautics and Astronautics, 1995.
- [2] Special Issue on Global Positioning System, *Proceedings of the IEEE*, Jan. 1999.
- [3] S. Haykin, *Adaptive Filter Theory*, 3rd ed. Englewood Cliffs, NJ: Prentice Hall, 1991.
- [4] D. A. Pados and S. N. Batalama, "Joint Space-Time Auxiliary-Vector Filtering for DS-CDMA Systems with Antenna Arrays," *IEEE Trans. Commun.*, vol. 47, pp. 1406-1415, Sept. 1999.
- [5] D. A. Pados and G. N. Karystinos, "An Iterative Algorithm for the Computation of the MVDR Filter," *IEEE Trans. Signal Process.*, submitted.

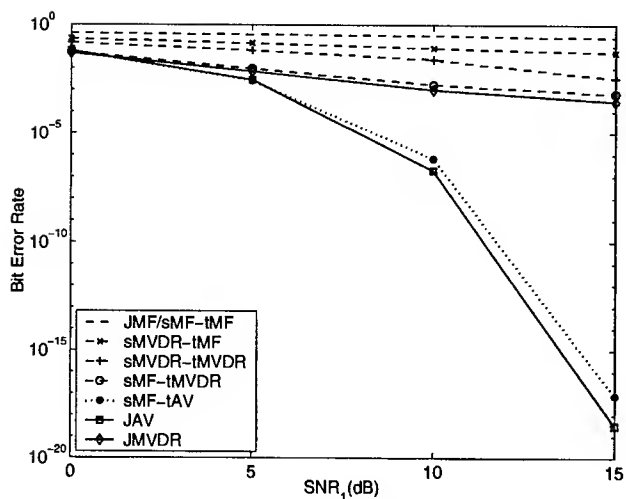


Fig. 1: Bit-Error-Rate as a function of the SNR of the user of interest.

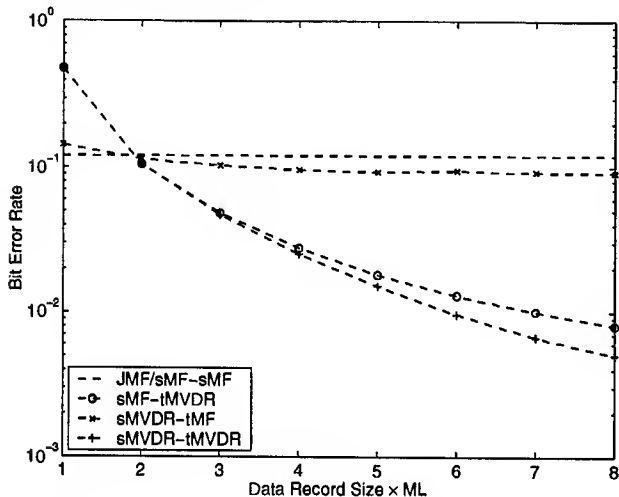


Fig. 2: Bit-Error-Rate as a function of the data record size for disjoint MF and MVDR-type filters.

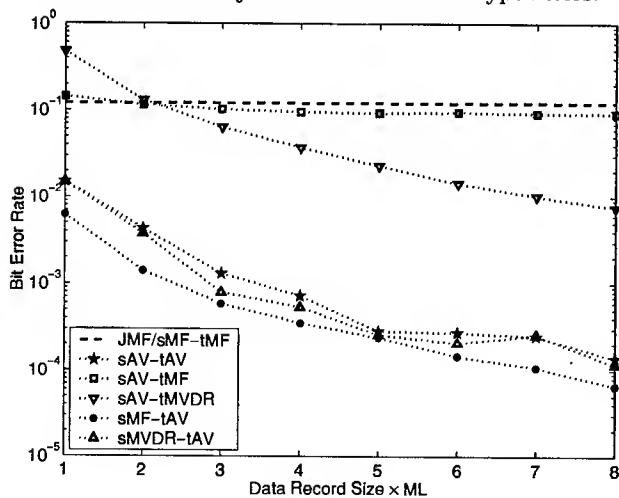


Fig. 3: Bit-Error-Rate as a function of the data record size for disjoint AV configurations.

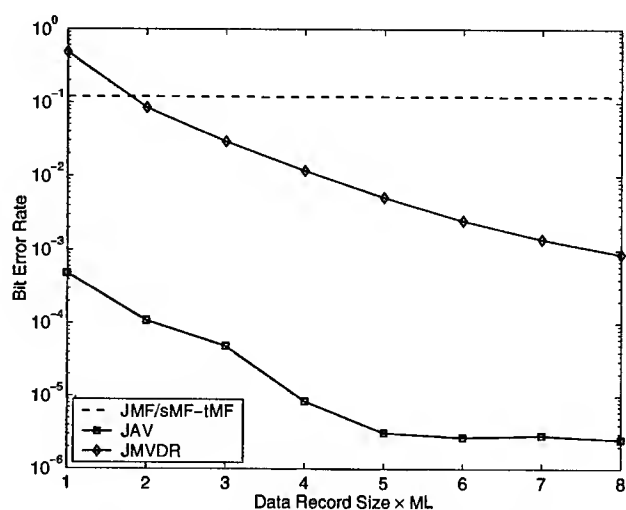


Fig. 4: Bit-Error-Rate as a function of the data record size for joint space-time configurations.

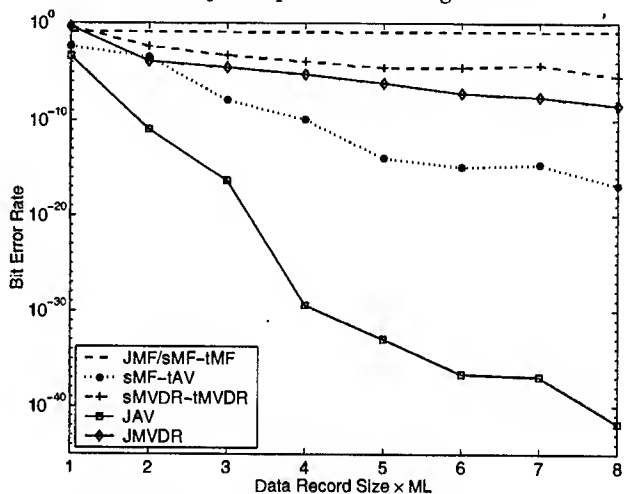


Fig. 5: Bit-Error-Rate as a function of the data record size after combining.

# SUBSPACE PROJECTION TECHNIQUES FOR ANTI-FM JAMMING GPS RECEIVERS

Liang Zhao<sup>†</sup>, Moeness G. Amin<sup>†</sup>, and Alan R. Lindsey<sup>‡</sup>

<sup>†</sup> Department of Electrical and Computer Engineering,  
Villanova University, Villanova, PA 19085, USA

E-mail: liang@ece.vill.edu, moeness@ece.vill.edu

<sup>‡</sup> Air Force Research Laboratory / IFGC  
525 Brooks Road, Rome, NY 13441, USA

E-mail: Alan.Lindsey@rl.af.mil

## ABSTRACT

*This paper applies subspace projection techniques as a pre-correlation signal processing method for the FM interference suppressions in GPS receivers. The FM jammers are instantaneous narrowband and have clear time-frequency (t-f) signatures that are distinct from the GPS C/A spread spectrum code. In the proposed technique, the instantaneous frequency (IF) of the jammer is estimated and used to construct a rotated signal space in which the jammer occupies one dimension. The anti-jamming system is implemented by projecting the received sequence onto the jammer-free subspace. This paper focuses on the characteristics of the GPS C/A code and derives the signal to interference and noise ratio (SINR) of the GPS receivers implementing the subspace projection techniques.*

## 1. INTRODUCTION

The Global Positioning System (GPS) is a satellite-based, worldwide, all-weather navigation and timing system [1]. The ever-increasing reliance on GPS for navigation and guidance has created a growing awareness of the need for adequate protection against both unintentional and intentional interference. Jamming is a procedure that attempts to block reception of the desired signal by the intended receiver. In general terms, it is high power signal that occupies the same frequency as the desired signal, making reception by the intended receiver difficult or impossible. Designers of military as well as commercial communication systems have, through the years, developed numerous anti-jamming techniques to counter these threats. As these techniques become effective for interference removal and mitigation, jammers themselves have become smarter and more sophisticated, and generate signals, which are difficult to combat.

The GPS system employs BPSK-modulated direct sequence spread spectrum (DSSS) signals. The DSSS systems

are implicitly able to provide a certain degree of protection against intentional or non-intentional jammers. However, in many cases, the jammer may be much stronger than the GPS signal, and the spreading gain might be insufficient to decode the useful data reliably [2]. There are several methods that have been proposed for interference suppression in DSSS communications [3, 4, 5]. The recent development of the bilinear time-frequency distributions (TFDs) for improved signal power localization in the time-frequency plane has motivated several new effective approaches, based on instantaneous frequency (IF) estimation, for non-stationary interference excisions [6]. One of the important IF-based interference rejection techniques uses the jammer IF to construct a time-varying excision notch filter that effectively removes the interference [7]. However, this notch filtering excision technique causes significant distortions to the desired signal, leading to undesired receiver performance.

Recently, subspace projection techniques, which are also based on IF estimation, have been devised for non-stationary FM interference excision in DSSS communications [8]. The techniques assume clear jammer time-frequency signatures and rely on the distinct differences in the localization properties between the jammer and the spread spectrum signals. The jammer instantaneous frequency, whether provided by the time-frequency distributions or any other IF estimator, is used to form an interference subspace. Projection can then be performed to excise the jammer from the incoming signal prior to correlation with the receiver PN sequence. The result is improved receiver SINR and reduced BERs.

In this paper, we apply the subspace projection techniques as a pre-correlation signal processing method to the FM interference suppression in GPS receivers. The GPS receiver and signal structure impose new constraints on the problem since the spreading code from each satellite is known and periodic within one navigation data symbol. This structure and the signal model are reviewed in Section 2. In Section 3, we depict the received GPS signal properties in time-frequency domain. The SINR of the GPS receiver implementing the subspace projection techniques

The work of Dr. M. Amin and A. Lindsey is supported by the Air Force Research Lab., Grant No. F30602-99-2-0504

is derived in Section 4, which shows improved performance in strong interference environments.

## 2. SIGNAL MODEL

GPS employs BPSK-modulated DSSS signals. The navigation data is transmitted at a symbol rate of 50 bps. It is spread by a coarse acquisition (C/A) code and a precision (P) code. The C/A code is a Gold sequence with a chip rate of 1.023 MHz and a period of 1023 chips, i.e. its period is 1ms, and there are 20 periods within one data symbol. The P code is a pseudorandom code at the rate of 10.23 MHz and with a period of 1 week. These two spreading codes are multiplexed in quadrature phases. Figure 1 shows the signal structure. The carrier L1 is modulated by both C/A code and P code, whereas the carrier L2 is only modulated by P code. In this paper, we will mainly address the problem of anti-jamming for the C/A code, for which the peak power spectral density exceeds that of the P code by about 13 dB [1]. The transmitted GPS signal is also very weak with Jammer-to-Signal Ratio (JSR) often larger than 40 dB and Signal-to-Noise Ratio (SNR) in the range -14 to -20 dB [2, 9]. Due to the high JSR, the FM jammer often has a clear signature in the time-frequency domain as shown in Section 3. As the P code is very weak compared to the C/A code, noise and jammer, we can ignore its presence in our analysis.

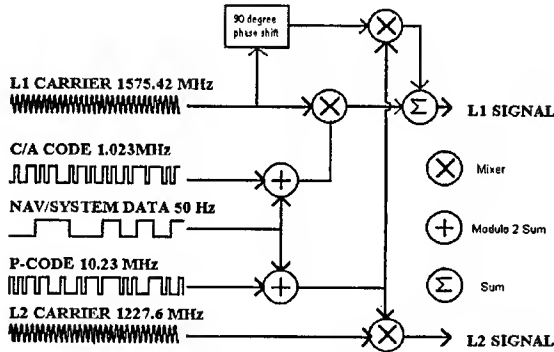


Figure 1: The GPS signal structure.

The BPSK-modulated DSSS signal may be expressed as

$$s(t) = \sum_i I_i b_i(t - iT_b) \quad I_i \in \{-1, 1\} \forall i \quad (1)$$

where  $I_i$  represents the binary information sequence and  $T_b$  is the bit interval, which is 20ms in the case of GPS system. The  $i^{th}$  binary information bit,  $b_i(t)$  is further decomposed as a superposition of  $L$  spreading codes,  $p(n)$ , pulse shaped by a unit-energy function,  $q(t)$ , of duration of  $\tau_c$ , which is 1/1023 ms in the case of C/A code. Accordingly,

$$b_i(t) = \sum_{n=1}^L p(n) q(t - n\tau_c) \quad (2)$$

The signal for one data bit at the receiver, after demodulation, and sampling at chip rate, becomes

$$x(n) = p(n) + w(n) + j(n) \quad 1 \leq n \leq L \quad (3)$$

where  $p(n)$  is the chip sequence,  $w(n)$  is the white noise, and  $j(n)$  is the interfering signal. The above equation can be written in the vector form

$$\mathbf{x} = \mathbf{p} + \mathbf{w} + \mathbf{j} \quad (4)$$

where

$$\begin{aligned} \mathbf{x} &= [x(1) \ x(2) \ x(3) \ \cdots \ x(L)]^T, \\ \mathbf{p} &= [p(1) \ p(2) \ p(3) \ \cdots \ p(L)]^T, \\ \mathbf{w} &= [w(1) \ w(2) \ w(3) \ \cdots \ w(L)]^T, \\ \mathbf{j} &= [j(1) \ j(2) \ j(3) \ \cdots \ j(L)]^T \end{aligned}$$

All vectors are of dimension  $L \times 1$ , and 'T' denotes vector or matrix transposition. It should be noted that the P vector is real, whereas all other vectors in the above equation have complex entries.

## 3. PERIODIC SIGNAL PLUS JAMMER IN THE TIME-FREQUENCY DOMAIN

For GPS C/A code, the PN sequence is periodic. The PN code of length 1023 repeats itself 20 times within one symbol of the 50 bps navigation data. Consequently, it is no longer of a continuous spectrum in the frequency domain, but rather of spectral lines. The case is the same for periodic jammers. Figure 2 and Figure 3 show the effect of periodicity of the signal and the jammer on their respective power distribution over time and frequency, using Wigner-Ville distribution. In both figures, a PN sequence of length 32 samples that repeats 8 times is used. A non-periodic chirp jammer of a 50dB JSR (jammer-to-signal ratio) is added in Figure 2. A periodic chirp jammer of 50 dB JSR with the same period as the C/A code is included in Figure 3. We note that the chosen value of 50dB JSR has a practical significance. The spread spectrum systems in a typical GPS C/A code receiver can tolerate a narrowband interference of approximately 40 dB JSR without interference mitigation processing. However, field tests show that jammer strength often exceeds that number due to the weakness of the signal. SNR in both figures are -20dB, which is also close to its practical value [2, 9]. Due to high JSR, the jammer is dominant in both figures. From Figure 3, it is clear that the periodicity of the jammer brings more difficulty to IF estimation than the non-periodic jammers. This problem can be solved by applying a short data window when using Wigner-Ville distribution. Note that the window length should be less than the jammer period. Figure 4 shows the result of applying a window of length 31 to the same data used in Fig. 3. It is evident from the Fig. 4 that the horizontal discrete harmonic lines have disappeared.

## 4. GPS ANTI-JAMMING USING PROJECTION TECHNIQUES

The concept of subspace projection for instantaneously narrowband jammer suppression is to remove the jammer com-

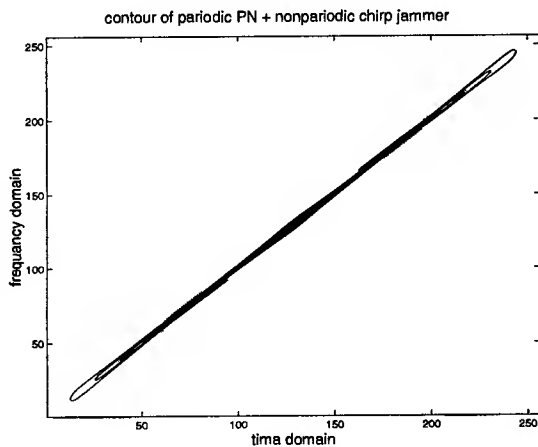


Figure 2: Periodic signal corrupted by a non-periodic jammer in time-frequency domain

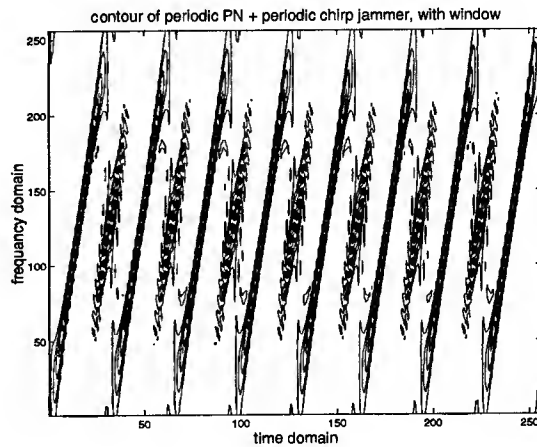


Figure 4: Periodic signal corrupted by a periodic jammer in time-frequency domain (with window)

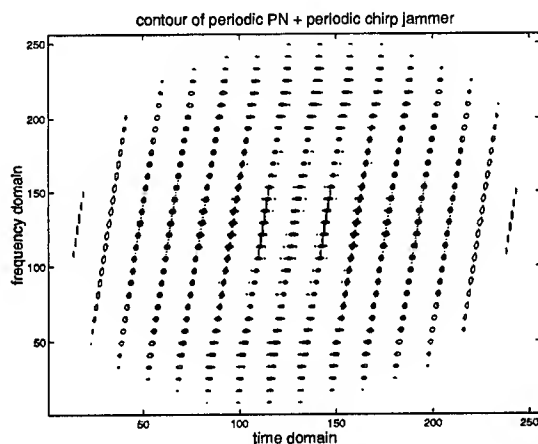


Figure 3: Periodic signal corrupted by a periodic jammer in time-frequency domain

ponents from the received data by projecting it onto the subspace that is orthogonal to the jammer subspace, as illustrated in Fig. 5.

Once the instantaneous frequency (IF) of the non-stationary jammer is estimated from the time-frequency domain, or by using any other IF estimator [10, 11, 12, 13], the interference signal vector  $\mathbf{j}$  in (4) can be constructed, up to ambiguity in phase and possibly in amplitude. In the proposed interference excision approach, the data vector is partitioned into  $Q$  blocks, each of length  $P$ , i.e.  $L=PQ$ . For the GPS C/A code,  $Q=20$ ,  $P=1023$ , and all  $Q$  blocks are identical, i.e., the signal PN sequence is periodic. Block-processing provides the flexibility to discard the portions of the data bit, over which there are significant errors in the IF estimates. The orthogonal projection method makes use of the fact that, in each block, the jammer has a one-dimensional subspace  $\mathcal{J}$  in the  $P$ -dimensional space  $\mathcal{V}$ , which is spanned by the received data vector. The interference can

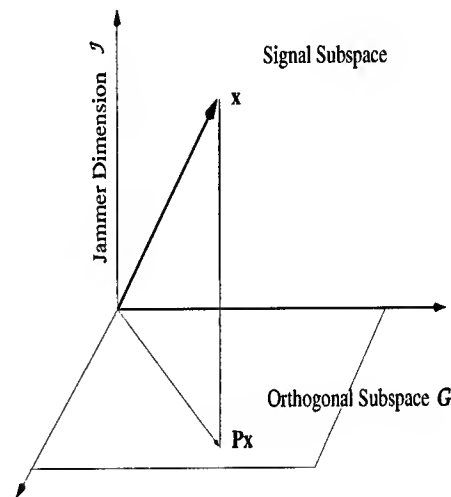


Figure 5: Jammer excision by subspace projection

be removed from each block by projecting the received data on the corresponding orthogonal subspace  $\mathcal{G}$  of the interference subspace  $\mathcal{J}$ . The subspace  $\mathcal{J}$  is estimated using the IF information. The projection matrix for the  $k^{th}$  block is given by

$$\mathbf{V}_k = \mathbf{I} - \mathbf{u}_k \mathbf{u}_k^H \quad (5)$$

The vector  $\mathbf{u}_k$  is the unit norm basis vector in the direction of the interference vector of the  $k^{th}$  block, and 'H' denotes vector or matrix Hermitian. Since the FM jammer signals are uniquely characterized by their IFs, the  $i^{th}$  FM jammer in the  $k^{th}$  block can be expressed as

$$u_k(i) = \frac{1}{\sqrt{P}} \exp[j\phi_k(i)] \quad (6)$$

The result of the projection over the  $k^{th}$  data block is

$$\tilde{\mathbf{x}}_k = \mathbf{V}_k \mathbf{x}_k \quad (7)$$

where  $\mathbf{x}_k$  is the input data vector. Using the three different components that make up the input vector in (4), the output of the projection filter  $\mathbf{V}_k$  can be written as

$$\bar{\mathbf{x}}_k = \mathbf{V}_k [\mathbf{p}_k + \mathbf{w}_k + \mathbf{j}_k] \quad (8)$$

The noise is assumed to be complex white Gaussian with zero-mean,

$$E[w(n)] = 0, E[w(n)^* w(n+l)] = \sigma^2 \delta(l), \forall l \quad (9)$$

Since we assume total interference excision through the projection operation, then

$$\mathbf{V}_k \mathbf{j}_k = 0, \quad \bar{\mathbf{x}}_k = \mathbf{V}_k \mathbf{p}_k + \mathbf{V}_k \mathbf{w}_k \quad (10)$$

The decision variable  $y_r$  is the real part of  $y$  that is obtained by correlating the filter output  $\bar{\mathbf{x}}_k$  with the corresponding  $k^{th}$  block of the receiver PN sequence and summing the results over the  $K$  blocks. That is,

$$y = \sum_{k=0}^{K-1} \bar{\mathbf{x}}_k^H \mathbf{p}_k \quad (11)$$

Since the PN code is periodic, we can strip off the subscript  $k$  in  $\mathbf{p}_k$ . The above variable can be written in terms of the constituent signals as

$$y = \sum_{k=0}^{Q-1} \mathbf{p}^T \mathbf{V}_k \mathbf{p} + \sum_{k=0}^{Q-1} \mathbf{w}^H \mathbf{V}_k \mathbf{p} \triangleq y_1 + y_2 \quad (12)$$

where  $y_1$  and  $y_2$  are the contributions of the PN and noise sequences to the decision variable, respectively. In [8],  $y_1$  is considered as a random variable. However, in GPS system, due to the fact that each satellite is assigned a fixed Gold code [1], and that the Gold code is the same for every navigation data symbol,  $y_1$  can no longer be treated as a random variable, but rather a deterministic value. This is a key difference between the GPS system and other spread spectrum systems. The value of  $y_1$  is given by

$$\begin{aligned} y_1 &= \sum_{k=0}^{Q-1} \mathbf{p}^T \mathbf{V}_k \mathbf{p} \\ &= \sum_{k=0}^{Q-1} \mathbf{p}^T (\mathbf{I} - \mathbf{u}_k \mathbf{u}_k^H) \mathbf{p} \\ &= \sum_{k=0}^{Q-1} (\mathbf{p}^T \mathbf{p} - \mathbf{p}^T \mathbf{u}_k \mathbf{u}_k^H \mathbf{p}) \\ &= QP - \sum_{k=0}^{Q-1} (\mathbf{p}^T \mathbf{u}_k \mathbf{u}_k^H \mathbf{p}) \end{aligned} \quad (13)$$

Define

$$\beta_k = \frac{\mathbf{p}^T \mathbf{u}_k}{\sqrt{P}} \quad (14)$$

as the correlation coefficient between the PN sequence vector  $\mathbf{p}$  and the jammer vector  $\mathbf{u}$ .  $\beta_k$  reflects the the component of the signal that is in the jammer subspace, and represents the degree of resemblance between the signal

sequence and the jammer sequence. Since the signal is a PN sequence, and the jammer is a non-stationary FM signal, the correlation coefficient is typically very small. With the above definition,  $y_1$  can be expressed as

$$y_1 = P(Q - \sum_{k=0}^{Q-1} |\beta_k|^2) \quad (15)$$

From (15), it is clear that  $y_1$  is a real value, which is the result of the fact that the projection matrix  $\mathbf{V}$  is Hermitian. With the assumptions in (9),  $y_2$  is complex white Gaussian with zero-mean. Therefore,

$$\begin{aligned} \sigma_{y_2}^2 &= E[|y_2|^2] \\ &= E\left[\left(\sum_{k=0}^{Q-1} \mathbf{w}^H \mathbf{V}_k \mathbf{p}\right)^H \left(\sum_{l=0}^{Q-1} \mathbf{w}^H \mathbf{V}_l \mathbf{p}\right)\right] \\ &= \sum_{k=0}^{Q-1} \sum_{l=0}^{Q-1} \mathbf{p}^T \mathbf{V}_k E[\mathbf{w}_k \mathbf{w}_l^H] \mathbf{V}_l \mathbf{p} \\ &= \sum_{k=0}^{Q-1} \mathbf{p}^T \mathbf{V}_k E[\mathbf{w}_k \mathbf{w}_k^H] \mathbf{V}_k \mathbf{p} \\ &= \sigma^2 \sum_{k=0}^{Q-1} \mathbf{p}^T \mathbf{V}_k \mathbf{V}_k \mathbf{p} \\ &= \sigma^2 \sum_{k=0}^{Q-1} \mathbf{p}^T \mathbf{V}_k \mathbf{p} = \sigma^2 y_1 \end{aligned} \quad (16)$$

the above equations make use of the noise assumptions in (9) and the properties of the projection matrix. The decision variable  $y_r$  is the real part of  $y$ . Consequently,  $y_r$  is given by

$$y_r = y_1 + \text{Re}\{y_2\} \quad (17)$$

where  $\text{Re}\{y_2\}$  denotes the real part of  $y_2$ .  $\text{Re}\{y_2\}$  is real white Gaussian with zero-mean and variance  $\frac{1}{2}\sigma_{y_2}^2$ . Therefore, the SINR is

$$\begin{aligned} \text{SINR} &= \frac{y_1^2}{\text{var}\{\text{Re}\{y_2\}\}} \\ &= \frac{y_1^2}{\frac{1}{2}\sigma_{y_2}^2} = \frac{2y_1}{\sigma^2} \\ &= \frac{2P(Q - \sum_{k=0}^{Q-1} |\beta_k|^2)}{\sigma^2} \end{aligned} \quad (18)$$

In the absence of jammers, no excision is necessary, and the SINR(SNR) of the receiver output will become  $2PQ/\sigma^2$ , which represents the upper bound for the anti-jamming performance. Clearly,  $\frac{2P \sum_{k=0}^{Q-1} |\beta_k|^2}{\sigma^2}$  is the reduction in the receiver performance caused by the proposed jammer suppression techniques. It reflects the energy of the power of the signal component that is in the jammer subspace. If the jammer and spread spectrum signals are orthogonal, i.e., their correlation coefficient  $|\beta| = 0$ , then interference suppression is achieved with no loss in performance. However, as stated above, in the general case,  $\beta_k$  is often very small, so the projection technique can excise FM jammers



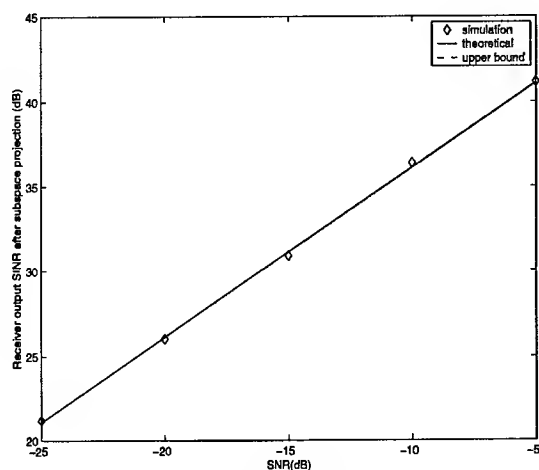


Figure 6: Receiver SINR vs SNR.

effectively with only very insignificant signal loss. The lower bound of SINR is zero and corresponds to  $|\beta| = 1$ . This case requires the jammer to assume the C/A code, i.e., identical and synchronous with actual one. Figure 6 depicts the theoretical SINR in (18), its upper bound, and estimated values using computer simulation. The SNR assumes five different values [-25, -20, -15, -10, -5] dB. In this figure, the signal is the Gold code of satellite SV#1, and the jammer is a periodic chirp FM signal with frequency 0–0.5 and has the same period as the C/A code. For this case, the correlation coefficient  $\beta$  is very small,  $|\beta| = 0.0387$ . JSR used in the computer simulation is set to 50dB. Due to the large computation involved, we have used 1000 realizations for each SNR value. Figure 6 demonstrates that the theoretical value of SINRs is almost the same as the upper bound and both are very close to the simulation result. In the simulation as well as in the derivation of equation (18), we have assumed exact knowledge of the jammer IF. Inaccuracies in the IF estimation will have an effect on the receiver performance [8].

## 5. CONCLUSIONS

GPS receivers are vulnerable to strong interferences. In this paper, subspace projection techniques are adapted for the anti-FM jamming GPS receiver. These techniques are based on IF estimation of the jammer signal, which can be easily achieved, providing that the C/A code and the jammer have distinct time-frequency signatures. The IF information is used to construct the FM interference subspace which, because of signal nonstationarities, is otherwise difficult to obtain. Due to the characteristic of the GPS spread spectrum signal structure and the fact that the C/A codes are fixed for the different satellites and known to all, the analysis of the receiver SINR becomes different from common spread spectrum systems. The theoretical and simulation results suggest that the subspace projection techniques can effectively excise FM jammers for GPS receivers with insignificant loss in the spreading gain.

## 6. REFERENCES

- [1] B. W. Parkinson and J. J. Spilker Jr. eds, *Global Positioning System: Theory and Applications*, American Institute of Aeronautics and Astronautics, 1996.
- [2] M. S. Braasch and A. J. Van Dierendonck, "GPS receiver architectures and measurements," *Proceedings of the IEEE*, vol. 87, no. 1, pp. 48–64, Jan. 1999.
- [3] L. B. Milstein, "Interference rejection techniques in spread spectrum communications," *Proceedings of the IEEE*, vol. 76, no. 6, pp. 657–671, Jun. 1988.
- [4] H. V. Poor and L. A. Rusch, "Narrowband interference suppression in spread spectrum CDMA," *IEEE Personal Communications Magazine*, third quarter, pp. 14–27, 1994.
- [5] J. D. Laster and J. H. Reed, "Interference rejection in digital wireless communications," *IEEE Signal Processing Magazine*, pp. 37–62, May 1997.
- [6] M. G. Amin and A. N. Akansu, "Time-frequency for interference excision in spread-spectrum communications," in "Highlights of Signal Processing for Communications," *IEEE Signal Processing Magazine*, vol. 16, no. 2, pp. 33–34, March 1999.
- [7] M.G. Amin, C. Wang and A. Lindsey, "Optimum interference excision in spread spectrum communications using open loop adaptive filters," *IEEE Trans. on Signal Processing*, vol. 47, no. 7, pp. 1966–1976, July 1999.
- [8] M.G. Amin, R. S. Ramineni and A. Lindsey, "Interference excision in DSSS communication systems using projection techniques," submitted to *IEEE Trans. on Signal Processing*, Jan. 2000.
- [9] R. Jr. Landry, P. Mouyon and D. Lekaim, "Interference mitigation in spread spectrum systems by wavelet coefficients thresholding," *European Trans. on Telecommunications*, vol. 9, pp. 191–202 March-Apr. 2000.
- [10] B. Boashash, "Estimating and interpreting the instantaneous frequency of a signal," Parts 1 and 2, *Proceedings of the IEEE*, vol. 80, no. 12, Dec. 1990.
- [11] S. Kay, "A fast and accurate single frequency estimator," *IEEE Trans. on Acoustics, Speech, and Signal Processing*, vol. 37, no. 12, Dec. 1979.
- [12] L. Cohen, *Time-Frequency Analysis*, Prentice Hall, Englewood Cliffs, New Jersey, 1995.
- [13] L. White, "Adaptive tracking of frequency modulated signals using hidden markov models," *Workshop on Hidden Markov Models For Tracking, Wirrina Cove Resort*, Feb. 1992.

# FIXED-POINT HAAR-WAVELET-BASED ECHO CANCELLER

Miloš Doroslovački, Iftikharuddin Khan

Bogdan Kosanović

George Washington University  
801 22nd Street, N.W.  
Washington D.C. 20052  
doroslov@seas.gwu.edu, ifti@seas.gwu.edu

Telogy Networks, Texas Instruments  
20250 Century Boulevard  
Germantown, MD 20874  
bkosanovic@telogy.com

## ABSTRACT

Telephone line echo path impulse responses have been estimated using a Haar-wavelet-based adaptive filter. Estimation error of two percent has been tolerated. This has allowed structural simplification to the Haar-wavelet-based adaptive filter. Adaptive filter coefficients are then updated by using a wavelet-based LMS algorithm, modified for fixed-point arithmetic. The fixed-point wavelet echo canceller has been tested for white noise and colored noise input signals and has been compared with the fixed-point FIR echo canceller through simulations. The wavelet based canceller converges much faster than its FIR counterpart.<sup>1</sup>

## 1. INTRODUCTION

Characteristics of echo paths are time varying. The corresponding impulse responses may have long tails and change frequently in the beginning. The length of FIR adaptive filter depends upon the tail of echo path impulse response and can therefore be large. On the other hand, if the same estimate of the echo path impulse response can be given, but with a lesser number of coefficients, the computational complexity is reduced. In a previous work [1], Haar wavelets have been used to estimate the echo path impulse response with lesser number of coefficients compared to FIR adaptive filter.

Wavelets are scaled and translated copies of a particular window function called the mother wavelet [2]. The wavelet transform, like the Fourier transform, can be used to decompose a function in terms of a set of basis function. The set of scaled and translated copies of mother wavelet, together with a set of functions known as the scaling functions form the basis in wavelet transform. Thus an unknown discrete-time system can be represented as

$$h(t) = \sum_{(m,n) \in D} a_{mn} \psi_{mn}(t) \quad (1)$$

<sup>1</sup>This work was supported by the Telogy Networks, Texas Instrument Company, under contract AN 19672

where  $\psi_{mn}(t)$  belongs to a set  $D$  of discrete-time Haar wavelets and  $a_{mn}$  are representation coefficients.

Haar wavelets are the simplest of wavelet functions. They are discrete-time orthonormal sequences  $\psi_{mn}(t)$  defined by

$$\psi_{mn}(t) = \psi_{m0}(t - 2^m n) \quad (2)$$

where

$$\psi_{m0}(t) = \begin{cases} 2^{-m/2} & \text{for } 0 \leq t \leq 2^{m-1} - 1; \\ -2^{-m/2} & \text{for } 2^{m-1} - 1 \leq t \leq 2^m - 1; \\ 0 & \text{otherwise.} \end{cases} \quad (3)$$

The indices  $m = 1, 2, \dots$  and  $n = \dots, -1, 0, 1, \dots$  correspond to the scale and translation respectively. Filtering of a signal by Haar wavelets and Haar scaling functions can be calculated as [3]

$$X_m(t) = \frac{X_{m-1} + X_{m-1}(t - 2^{m-1})}{\sqrt{2}} \quad (4)$$

$$Y_m(t) = \frac{X_{m-1} - X_{m-1}(t - 2^{m-1})}{\sqrt{2}} \quad (5)$$

where  $X_0(t)$  is the input signal and  $X_m(t)$  and  $Y_m(t)$  are the outputs of the scaling and wavelet filter at level  $m$ . The index  $m$  identifies the level.

Practically, Haar wavelet filtering is implemented by using two filters. The average component (or scaling filtering) of the input signal is obtained by passing  $X_{m-1}(t)$  through a low-pass filter. The difference component (or wavelet filtering) is obtained by passing  $X_{m-1}(t)$  through a high-pass filter. The multiresolution analysis can then be obtained by running the Haar filtering again on the average component that was obtained from (4). This implies that another set of two filters is required. The two-filter bank can be cascaded until the desired level of wavelet decomposition is reached.

## 2. SIZE OF WAVELET FILTER BANK

The size of the wavelet filter bank depends on what type of echo path impulse response it has to estimate.

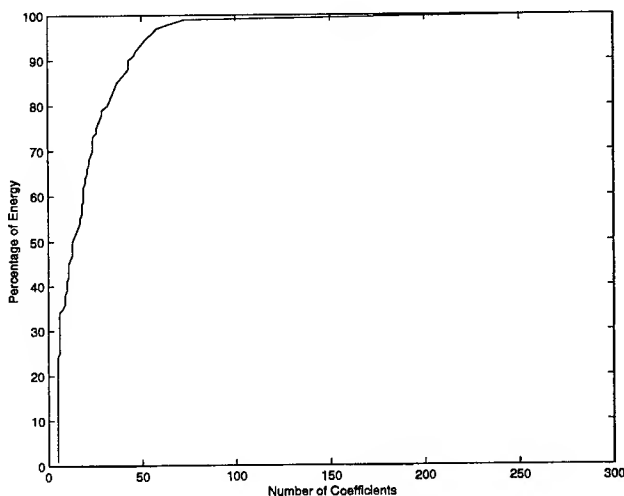


Figure 1: Energy corresponding to the number of coefficients

Once the maximum length of echo path impulse response,  $L_{max}$ , is known

$$L = \log_2(L_{max}) - 1$$

can be used to give the number of levels required in the wavelet filter bank. In order to simplify the complexity of wavelet adaptive filter the set  $\hat{D}$  instead of the set  $D$  is used in (1). One criterion to select  $\hat{D}$  can be to find the relationship between the energy of the hybrid and the wavelet and the scaling coefficients that are used to represent the hybrid. An algorithm is generated to calculate this relationship [3]. Twenty-one hybrids are then presented to this algorithm. The algorithm takes the Haar-wavelet transform of each hybrid, sorts the wavelet and scaling coefficients and gives the strongest coefficients that represent the required fraction of the total energy. These coefficients are referred as significant coefficients. The same operation is applied to all hybrids. The union of these significant coefficients is taken and used in the wavelet adaptive filter. The relationship between the energy and the number of wavelet coefficients is shown in Figure 1. In order to simplify the computational complexity, modeling error of two percent has been tolerated. The number of coefficients required to give estimation error of at most two percent, for every considered hybrid is 65. It has to be noted that these coefficients are not the first 65 coefficients but are distributed along scales and shifts.

### 3. HAAR WAVELET ADAPTIVE FILTER

Figure 2 shows complete wavelet filter bank that is used for adaptive estimation of the impulse response of a hybrid given in (1). This filter bank has a tree structure with each leaf of tree corresponding to a level of

Haar wavelet filtering. For instance, the filter  $H(\omega)$ , does the wavelet filtering at level 1. Similarly, the filter  $H(2\omega)$ , does the wavelet filtering at level 2. Figure 2 also shows delay elements at each level. The signal at the output of each delay element is referred as the shift of the signal at that level. Output at each delay element has an adaptive filter coefficient associated with it. The wavelet filtered signal present at the output of the high pass filter at a level and its shifted versions are then multiplied by their respective adaptive filter coefficients. The multiplication results are then summed up to give the output at that level. The output of the adaptive filter is then obtained by adding the outputs at each level. It has to be noted that the last level has two outputs - one coming from its wavelet filter and the other from its scaling filter.

#### 3.1. Floating Point Wavelet LMS

The floating point algorithm to adapt the filter coefficients is given as [4]

$$a_{mn}(t+1) = a_{mn}(t) + \mu_{mn} R_{mn}(t) e(t) \quad (6)$$

Here  $\mu_{mn}$  is the step size,  $e(t)$  the error between the desired signal and the output of the adaptive filter given as

$$z(t) = \sum_{(m,n) \in \hat{D}} R_{mn}(t) a_{mn}(t) \quad (7)$$

where  $\hat{D} \subseteq D$ . Using  $\hat{D}$  instead of  $D$  gives reduced order modeling.

#### 3.2. Fixed Point Wavelet LMS

The algorithm given in (6) can be modified for fixed point implementations. First, it has to be noted that the calculation of step size in (6) for each wavelet coefficient is not necessary as  $R_{mn}(t)$  is the same as  $R_{m0}(t) = Y_m(t)$  present at the output of high pass filter at level  $m$ , except for a delay. Therefore, the algorithm will function properly if the step size is calculated only once for each level as

$$\mu_{m0} = \frac{2c}{CP_m(t)} \quad (8)$$

where  $c$  and  $C$  are some constants and  $P_m(t)$  is the exponential-window time averaged power of a wavelet filter output. Exponential-window averaging of power of the wavelet filter output  $Y_m(t)$  is given as

$$\begin{aligned} P_m(t) &= (1 - \alpha)P_m(t-1) + \alpha Y_m^2(t) \\ &= P_m(t-1) + \alpha[Y_m^2(t) - P_m(t-1)] \end{aligned} \quad (9)$$

where  $\alpha = 2^{-k}$  and  $k \in \{0, 1, 2, \dots, 30\}$ .

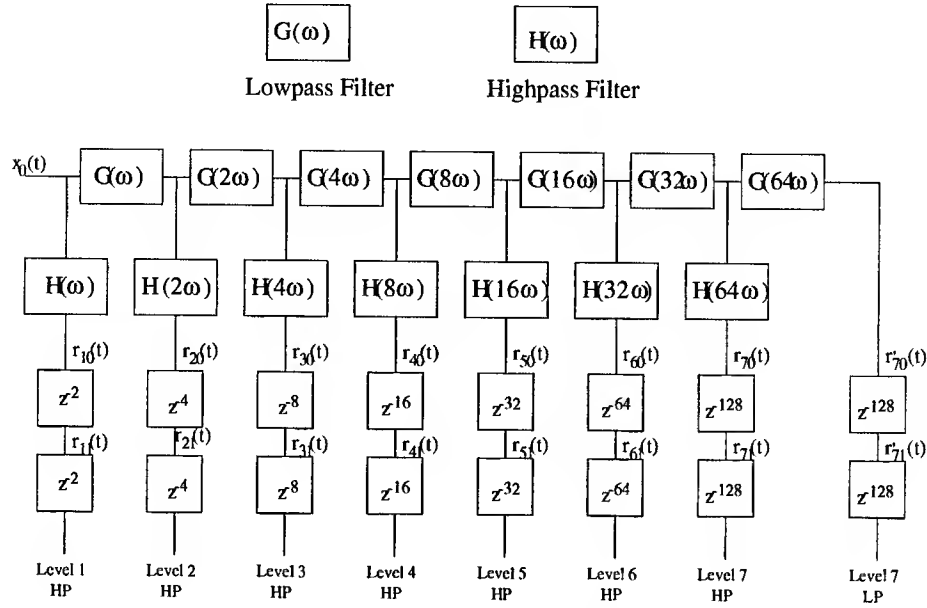


Figure 2: Wavelet Filter Bank

Division by square root of two is avoided in (4) and (5) by applying filtering as

$$x_m(t) = \frac{x_{m-1}(t) + x_{m-1}(t - 2^{m-1})}{2} \quad (10)$$

$$y_m(t) = \frac{x_{m-1}(t) - x_{m-1}(t - 2^{m-1})}{2} \quad (11)$$

The relationship with the previous Haar wavelet and scaling filter outputs is given as

$$X_m(t) = 2^{m/2} x_m(t) \quad (12)$$

$$Y_m(t) = 2^{m/2} y_m(t). \quad (13)$$

The output of modified filters,  $r_m(t)$ , is related to  $R_m(t)$  as

$$R_m(t) = 2^{m/2} r_m(t) \quad (14)$$

and the exponential-window time averaged power of  $y_m(t)$ ,  $p_m(t)$ , to  $P_m(t)$  as

$$P_m(t) = (2^{m/2})^2 p_m(t) \quad (15)$$

Substituting (14) in (6) and multiplying by  $2^{m/2}$  gives

$$\dot{a}_{mn}(t+1) = \dot{a}_{mn}(t) + \mu_{m0}(t) 2^m r_{mn}(t) e(t). \quad (16)$$

The step size can now be calculated as

$$\mu_{m0}(t) = \frac{2c}{C 2^m p_m(t)}. \quad (17)$$

The division operation in the step size calculation can be approximated by

$$\frac{2c}{C 2^m p_m(t)} \approx 2^{-w_m(t) - \beta - m} \quad (18)$$

where  $w_m(t) = \lceil \log_2 p_m(t) \rceil$  and  $\beta$  is some natural number.

An upper bound on adaptive filter coefficients given in (6) is

$$\begin{aligned} |a_{mn}| &= |\mathbf{h}^T \mathbf{W}_{mn}| = |\sum_k h(k) W_{mn}(k)| \\ &\leq \sqrt{[\sum_k h^2(k)] [\sum_k W_{mn}^2(k)]} \\ &= \sqrt{[\sum_k h^2(k)]} < 1 \end{aligned} \quad (19)$$

where  $\mathbf{h}$  is a vector containing hybrid's impulse response and  $\mathbf{W}_{mn}$  is a vector containing an orthonormal wavelet. In order to represent adaptive filter coefficients using the full sixteen bit binary number range, (16) needs to be multiplied by  $2^{15 - \lceil m/2 \rceil}$ . This results in

$$\ddot{a}_{mn}(t+1) = \ddot{a}_{mn}(t) + \frac{2c}{C p_m(t)} 2^{15 - \lceil m/2 \rceil} r_{mn}(t) e(t) \quad (20)$$

where  $\ddot{a}_m(t) = 2^{15 - \lceil m/2 \rceil} \dot{a}_m(t)$ . The fixed-point wavelet LMS algorithm can now be written as

$$\ddot{a}_{mn}(t+1) = \ddot{a}_{mn}(t) + 2^{-[w_m(t) + \beta - 15]} r_{mn}(t) e(t) 2^{-m/2}. \quad (21)$$

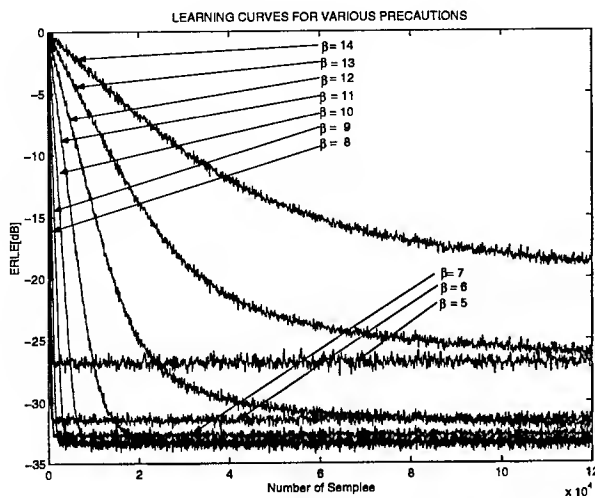


Figure 3: Learning curve with various precautions for wavelet LMS

Precaution $\beta$	Steady State Error(dB)	Convergence Rate(dB/sec)
5	-26.88	-270
6	-31.4471	-394
7	-32.69	-284
8	-33.1980	-160
9	-33.427	-83
10	-33.5345	-43
11	-33.59	-22
12	-33.6(est)	-11
13	-33.6(est)	-5
14	-33.6(est)	-2

Table 1: Steady state error and rate of convergence for wavelet LMS

#### 4. PERFORMANCE OF FIXED-POINT WAVELET LMS ALGORITHM

The fixed-point wavelet LMS depends on three parameters:  $\alpha$ , which governs the size of window for calculation of power;  $\beta$ , the precaution in step size; and  $P_m(0)$ , the initial value of power in (9). Through simulations it was noted that changing the parameters,  $\alpha$  and  $P_m(0)$ , have no major effect on the performance. Parameter  $\beta$  is a precaution factor, as it is inversely related to the step size. Therefore, it was observed that increasing the value of  $\beta$  reduces the speed of convergence.

##### 4.1. White Noise

Fixed-point wavelet LMS can be compared with fixed-point FIR LMS, using Figure 3 and Figure 4, for white noise far-end signal with power  $10^8$  and near-end noise power level equal to -35dB with respect to the output

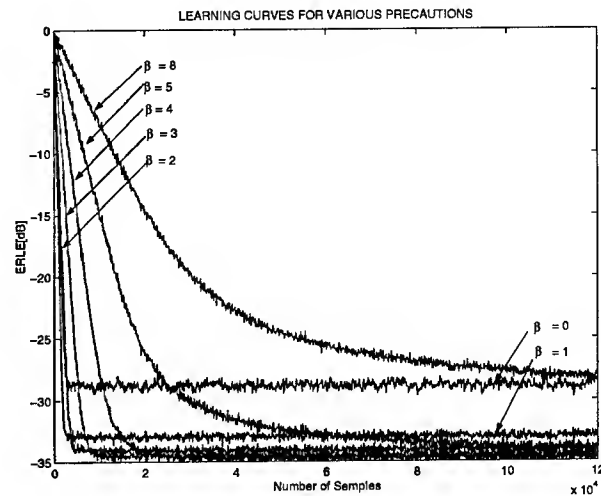


Figure 4: Learning curves for different precautions for FIR LMS

Precaution $\beta$	Steady State Error(dB)	Convergence Rate(dB/sec)
0	-28.86	-92
1	-32.9	-118
2	-34	-75
3	-34.5	-38
4	-34.7	-18
5	-35(est)	-9
6	-35(est)	-5

Table 2: Steady state error and rate of convergence for FIR LMS

of hybrid. The near-end signal is introduced because of two reasons: (a) to provide possibility to match steady state errors for wavelet and FIR adaptive algorithms, and (b) to model more accurately the real life working environment. The plots are obtained after 10 Monte Carlo runs. Tab.1 and Tab.2 show the steady state error and rate of convergence for different precautions for wavelet LMS and FIR LMS, respectively. Using the values from these tables,  $\beta=1$  for FIR and  $\beta=7$  for wavelet LMS can be used to match the steady state error. Similarly,  $\beta=3$  for FIR and  $\beta=10$  for wavelet LMS can be used to match the convergence rate. Some steady state errors in the tables are predicted since it takes long time to reach the convergence. This is denoted in the tables by *est*.

##### 4.2. Colored Noise

Three kinds of colored noise are considered : low pass, high pass and band pass, as the far-end signal. The near-end noise is white and has a power level of -35 dB with respect to the output of hybrid. Butterworth fil-

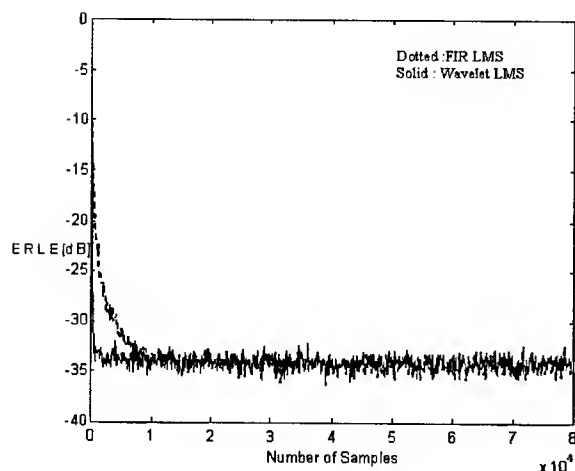


Figure 5: Low-pass colored noise with  $\beta = 3$  for FIR and  $\beta = 7$  for wavelet

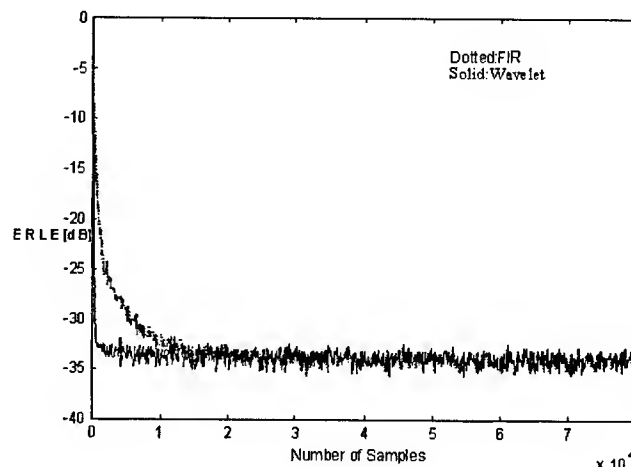


Figure 7: High-pass colored noise with  $\beta = 3$  for FIR and  $\beta = 7$  for wavelet

ters of order 18 with three different pass bands are used to generate the colored noise. For band pass noise, the filter has a pass band ranging from 1200 to 2800 Hz. For low pass noise, the filter has a cutoff frequency of 800 Hz. For high pass noise, the filter passes all the frequencies higher than 2800 Hz. The plots shown in Figure 5, Figure 6 and Figure 7 are for the case when the steady state error is matched directly for both algorithms. It can be seen that wavelet algorithm converges noticeably faster. Similar conclusions are obtained when the far-end input is voice, sinusoidal, and composite source signal [5], as it is documented in [3].

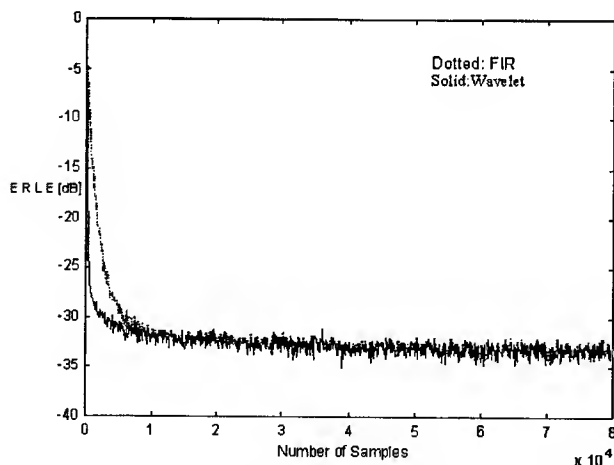


Figure 6: Bandpass colored noise with  $\beta = 3$  for FIR and  $\beta = 7$  for wavelet

## REFERENCES

- [1] M. Doroslovački and H. Fan, "On line identification of echo path impulse responses by Haar wavelet based adaptive filter," *Proc. IEEE Int. Conf. Acoust., Speech, Signal Processing*, vol. 2, pp. 1065-1068, May 1995.
- [2] S. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, 1998.
- [3] I. Khan, *Haar-wavelet-based echo canceller*, Master's thesis, Department of Electrical and Computer Engineering, George Washington University, Washington, DC 20052, January 2000.
- [4] M. Doroslovački and H. Fan, "Wavelet based linear system modeling and adaptive filtering," *IEEE Trans. Signal Processing*, vol. 44, no.5, pp. 1156-1167, May 1996.
- [5] *ITU-T Recommendation G.168: Digital Network Echo Cancellers*, International Telecommunication Union, Geneva, Switzerland, April 1997.

# WAVELET-POLYSPECTRA: ANALYSIS OF NON-STATIONARY AND NON-GAUSSIAN/NON-LINEAR SIGNALS

Yngvar Larsen and Alfred Hanssen

University of Tromsø, Physics Department  
Electrical Engineering Group, N-9037 Tromsø, Norway  
E-mail: yngvarl@phys.uit.no, alfred@phys.uit.no

## ABSTRACT

We present a stringent definition of higher-order evolutionary spectra. On this basis, we define wavelet-polyspectral densities as a way of dealing with non-stationarities in higher-order statistics. We propose a simple wavelet-polyspectral estimator, and we discuss its statistical properties. The proposed wavelet-polyspectral analysis tool is demonstrated by a numerical example. It is concluded that the wavelet-polyspectra have desirable properties for the analysis of data that are simultaneously non-stationary and non-Gaussian/non-linear.

## 1. INTRODUCTION

Non-stationary phenomena have traditionally been studied by (naive) spectrogram techniques [1, 2]. These methods suffer from poor statistical behavior and poor resolution in time and/or frequency. Various non-linear time-frequency methods (e.g. Cohen's class [1] and the Wigner-Ville distribution [2]) have been suggested to mend some of the mentioned weaknesses, but the use of these is again non-trivial and may be hard to interpret in practice. The wavelet transform [2] is a linear transform method that has rapidly become a popular technique to quantify scale-time variations of time series.

For almost all known polyspectral estimators, it is a requirement that the data are stationary. This is among other things due to the fact that most definitions of polyspectral estimators contain Fourier-transforms of the data, which provide no time information in the transformed domain. Note however that polyspectral analysis has been combined with the Wigner-Ville formalism [3, 4] to cope with non-stationarities.

Recently, it has been suggested [5, 6] to combine the normalized third-order polyspectrum (the bicoherence) with the wavelet transform, forming the *wavelet-bicoherence*.

The main purpose of this paper is to introduce a precise definition of wavelet-polyspectra of general orders, to propose estimators and discuss the statistical properties of these. In addition, we demonstrate the proposed technique by a relevant numerical example.

## 2. HIGHER-ORDER SPECTRA OF NON-STATIONARY PROCESSES

In [7] the class of *oscillatory processes*, which admits a representation of the form

$$X(t) = \int_{-\infty}^{\infty} A_t(f) \exp(j2\pi ft) dZ(f) \quad (1)$$

with  $A_t(f) = A_t^*(-f)$ , is examined. Here  $Z(f)$  is an orthogonal process with  $E\{|dZ(f)|^2\} = d\mu_2(f)$ . The measure  $\mu_2(f)$  is then an analog of the integrated spectrum in the stationary case. For such processes we define the  $n$ th-order evolutionary spectrum  $C_n(f_1, \dots, f_{n-1}, t)$  with respect to a family  $\mathcal{F} = \{A_t(f) \exp(j2\pi ft)\}$  of oscillatory functions by

$$\begin{aligned} dC_t^n(f_1, \dots, f_n) &= A_t(f_1) \cdots A_t(f_{n-1}) A_t(f_n) \\ &\quad \cdot \text{Cum}[dZ(f_1), \dots, dZ(f_n)] \\ &= A_t(f_1) \cdots A_t(f_{n-1}) A_t(f_n) \\ &\quad \cdot \delta(f_1 + \dots + f_n) d\mu_n(f_1, \dots, f_{n-1}) \\ &= A_t(f_1) \cdots A_t(f_{n-1}) A_t^*(f) \\ &\quad \cdot \mu_n(f_1, \dots, f_{n-1}) df_1 \cdots df_{n-1} \\ &\triangleq C_n(f_1, \dots, f_{n-1}, t) df_1 \cdots df_{n-1} \end{aligned} \quad (2)$$

provided that the zero-mean increment process  $dZ(f)$  is at least stationary to order  $n$ , and that the measure  $d\mu_n(f_1, \dots, f_{n-1})$  is absolutely continuous with respect to the Lebesgue measure. Here  $f = f_1 + \dots + f_{n-1}$  and  $\text{Cum}[\cdot]$  denotes the cumulant sequence. Note that the evolutionary power spectrum defined in [7] is a special case of the above definition with  $n = 2$ .

### 3. WAVELET-POLYSPECTRA

#### 3.1. Definitions

The continuous wavelet transform (CWT) of a real valued signal  $x(t)$  is defined as [2]

$$W_\psi(t, a) \triangleq \frac{1}{\sqrt{a}} \int_{-\infty}^{\infty} x(t') \psi^* \left( \frac{t' - t}{a} \right) dt'. \quad (3)$$

Based on the CWT we propose to define the  $n$ th-order wavelet-polyspectrum with respect to a wavelet of the form  $\psi(t) = g(t) \exp(j\eta t)$  as

$$M_n^w(a_1, \dots, a_{n-1}; t_0) \triangleq \frac{1}{T} E \left\{ \int_{t_0 - T/2}^{t_0 + T/2} W_\psi(t, a_1) \cdots W_\psi(t, a_{n-1}) W_\psi^*(t, a) dt \right\}, \quad (4)$$

where the process  $X(t)$  is assumed to be stationary on the interval of integration  $[t_0 - T/2, t_0 + T/2]$  and

$$a^{-1} = a_1^{-1} + a_2^{-1} + \dots + a_{n-1}^{-1}. \quad (5)$$

The reason for this *inverse sum rule* will become clear in the following section.

#### 3.2. Properties

In this paper, we will constrain the wavelets to the class defined by

$$\psi(t) = g(t) \exp(j\eta t), \quad (6)$$

where  $g(t)$  is a real valued and symmetric window which has to be well localized in the frequency domain. The parameter  $\eta$  is chosen such that the Fourier transform of  $\psi(t)$  is essentially zero for  $f \leq 0$ . With this choice of wavelet there is a well defined relationship between frequency  $f$  and scale  $a$ . For a piecewise stationary process this relationship is given by  $a = \eta/(2\pi f)$  for  $f \neq 0$ . The CWT is therefore not defined for frequency  $f = 0$ .

The inverse relationship between  $f$  and  $a$  is the reason why we apply the inverse sum rule (5) in the definition of the wavelet-polyspectra instead of the sum rule used in ordinary polyspectra [8]. Using this relation and considering the limit of (4) as  $T \rightarrow \infty$  it can be shown [9] that the  $n$ th-order wavelet-polyspectrum of a real valued *stationary* process is related to the ordinary  $n$ th-order moment spectrum  $M_n$  by

$$M_n^w(a_1, \dots, a_{n-1}) = G_n * M_n(f_1, \dots, f_{n-1}). \quad (7)$$

Here  $*$  denotes  $(n-1)$ -dimensional convolution and  $G_n$  is a spectral window which is different for each frequency tuple  $(f_1, \dots, f_{n-1})$ . This window acts as a

constant-Q smoothing filter in the frequency domain, and its exact form depends on the choice of window  $g(t)$  in the wavelet (6).

### 4. WAVELET-POLYSPECTRAL ESTIMATION

#### 4.1. Estimators

A straightforward method of estimating the  $n$ th-order wavelet spectrum is to discretize the time in equation (4) and remove the expectation operator. Then if  $t' = n\Delta t$  and  $t = k\Delta t$ ;  $n, k = 0, 1, \dots, N-1$ , a time-discretized version of the wavelet transform in equation (3) becomes a suitable estimator for the CWT. Assuming  $\Delta t = 1$ , this becomes

$$\widehat{W}_\psi(k, f) = \sqrt{\frac{2\pi f}{\eta}} \sum_{n=0}^{N-1} x(n) \psi^* \left[ \frac{2\pi f}{\eta} (n - k) \right], \quad (8)$$

where we have used  $a = \eta/(2\pi f)$ . If we let  $T = 2K$ , an estimator for the  $n$ th-order wavelet spectrum at time  $t_0 = L$  is

$$\widehat{M}_n^w(f_1, \dots, f_{n-1}; L) = \frac{1}{2K+1} \sum_{k=L-K}^{L+K} \widehat{W}_\psi(k, f_1) \cdots \widehat{W}_\psi(k, f_{n-1}) \widehat{W}_\psi^*(k, f), \quad (9)$$

where  $f = f_1 + f_2 + \dots + f_{n-1}$ . We require that the process is stationary in the time interval  $[L-K, L+K]$ .

#### 4.2. Statistical properties

The statistical properties of the estimator in equation (9) are quite complicated. For simplicity we will only discuss some results for  $n = 2$  and  $n = 3$  assuming a zero mean process.

The expected value of the estimated second-order wavelet spectrum of a zero mean stationary process  $X(t)$  can be written as [9]

$$E \left\{ \widehat{M}_2^w(f) \right\} = \int_{-1/2}^{1/2} V_2(f - f'; f) S(f') df'. \quad (10)$$

Here  $S(f)$  is the true power spectrum and the spectral window  $V_2(f'; f)$  is given by

$$V_2(f'; f) = \frac{1}{N} \sum_{k=0}^{N-1} |G(f'; k, f)|^2, \quad (11)$$

where  $G(f'; k, f)$  is the discrete Fourier transform of a scaled and shifted version of the window  $g(t)$  in the



wavelet (6) given by

$$g_{k,f}(n) = \sqrt{\frac{2\pi f}{\eta}} g\left[\frac{2\pi f}{\eta}(n-k)\right]. \quad (12)$$

Notice that this window is *frequency dependent*. For a fixed frequency  $f$ , equation (10) is a convolution between the spectral window  $V_2(f'; f)$  and the true power spectrum. With a proper (frequency dependent) normalization, one can show that  $\widehat{M}_2^w(f)$  is an asymptotically unbiased estimate of the true power spectrum  $S(f)$ . Similarly, in the non-stationary case, a normalization yields an asymptotically unbiased estimate of the evolutionary power spectrum  $C_2(f, t)$  defined in equation (2).

The support of the window  $g_{k,f}[n]$  is shorter than  $N$ , the number of samples (except at low frequencies). Thus the estimator  $\widehat{M}_2^w(f)$  is equivalent to the Weighted Overlapped Segment Averaging (WOSA) spectral estimation method for a fixed frequency  $f$ , since we can write

$$\widehat{M}_2^w(f) = \frac{1}{N} \sum_{k=0}^{N-1} \left| \sum_{n=0}^{N-1} x(n) g_{k,f}(n) \exp(-j2\pi f n) \right|^2.$$

The WOSA spectral estimator is well known for being consistent [10]. As mentioned, the estimator in the equation above may be normalized to yield an unbiased power spectral estimator. The variance of this unbiased estimator for stationary Gaussian signals may be approximated by

$$\text{Var}\{\widehat{S}(f)\} \approx \frac{S^2(f)}{N} \left\{ 1 + 2 \sum_{\tau=1}^{N-1} \left(1 - \frac{\tau}{N}\right) \left| \frac{g_2(\tau; f)}{g_2(0; f)} \right|^2 \right\},$$

where  $S(f)$  is the true power spectrum and

$$g_2(\tau; f) = \frac{2\pi f}{\eta} \int_{-\infty}^{\infty} g\left[\frac{2\pi f}{\eta}t\right] g\left[\frac{2\pi f}{\eta}(t+\tau)\right] dt$$

is the correlation between overlapping windows. The variance of the periodogram  $\widehat{S}^p(f)$  under the same assumptions can be written as  $\text{Var}\{\widehat{S}^p(f)\} \approx S^2(f)$  (see e.g. [10]). We thus see that the variance is reduced by a factor  $N/\nu_N(f)$ , where

$$\nu_N(f) = 1 + 2 \sum_{\tau=1}^{N-1} \left(1 - \frac{\tau}{N}\right) \left| \frac{g_2(\tau; f)}{g_2(0; f)} \right|^2,$$

relative to a (possibly tapered) periodogram. This frequency dependent factor increases with frequency, since

the width of the correlation function  $g_2(\tau; f)$  between two overlapping windows decreases with frequency due to the constant-Q property of the CWT.

In the non-stationary case one can show that the expected value of  $\widehat{M}_2^w(f; t)$  at  $t = L$  is approximately given by [9]

$$E\{\widehat{M}_2^w(f; L)\} \approx \int_{-1/2}^{1/2} \tilde{V}_2(f - f'; f) C_2(f', L) df',$$

where the total spectral window  $\tilde{V}_2(f'; f)$  is given by

$$\tilde{V}_2(f'; f) = \frac{1}{2K+1} \sum_{k=L-K}^{L+K} |G(f'; k, f)|^2.$$

This shows that the estimator is approximately unbiased for the true evolutionary power spectrum  $C_2(f, t)$  with proper normalization. The variance of this estimator is essentially the same as in the stationary case, with  $N$  replaced by  $2K+1$ , the length of the interval where the process is assumed stationary. Obviously, the estimator is not consistent due to the nonstationary nature of the process.

The expected value of the estimated third-order wavelet spectrum of a zero mean stationary process  $X(t)$  can be written as [9]

$$E\{\widehat{M}_3^w(f_1, f_2)\} = \int_{-1/2}^{1/2} V_3(f_1 - f'_1, f_2 - f'_2; f_1, f_2) B(f'_1, f'_2) df'_1 df'_2. \quad (13)$$

Here  $B(f_1, f_2)$  is the true bispectrum and the total bispectral window  $V_3(f'_1, f'_2; f_1, f_2)$  is given by

$$V_3(f'_1, f'_2; f_1, f_2) = \frac{1}{N} \sum_{k=0}^{N-1} G(f'_1; k, f_1) G(f'_2; k, f_2) G^*(f'_1 + f'_2; k, f_1 + f_2). \quad (14)$$

The estimator may, as in the second-order case, be normalized to yield an asymptotically unbiased estimate [9]. Assuming a stationary Gaussian process, the variance of the normalized estimator may be written as

$$\text{Var}\{\widehat{B}(f_1, f_2)\} \approx \frac{S(f_1)S(f_2)S(f)}{N} \left[ \frac{g_2(0; f_1)g_2(0; f_2)g_2(0; f)}{|g_3(0, 0; f_1, f_2)|^2} + 2 \sum_{\tau=1}^{N-1} \left(1 - \frac{\tau}{N}\right) \frac{g_2(\tau; f_1)g_2(\tau; f_2)g_2(\tau; f)}{|g_3(0, 0; f_1, f_2)|^2} \right],$$

where  $g_3(\tau_1, \tau_2; f_1, f_2)$  is the triple correlation given by

$$g_3(\tau_1, \tau_2; f_1, f_2) = \left( \frac{2\pi f}{\eta} \right)^{\frac{3}{2}} \int_{-\infty}^{\infty} g \left[ \frac{2\pi f}{\eta} t \right] \cdot g \left[ \frac{2\pi f_1}{\eta} (t + \tau_1) \right] g \left[ \frac{2\pi f_2}{\eta} (t + \tau_2) \right] dt$$

and  $f = f_1 + f_2$ . The variance of the biperiodogram  $\widehat{B}^p(f_1, f_2)$  under the same assumptions can be written as  $\text{Var} \left\{ \widehat{B}^p(f_1, f_2) \right\} \approx NS(f_1)S(f_2)S(f)$  (see e.g. [10]). We thus see that the variance is reduced by a factor  $N^2/\nu_N(f_1, f_2)$ , where

$$\nu_N(f_1, f_2) = \frac{g_2(0; f_1)g_2(0; f_2)g_2(0; f)}{|g_3(0, 0; f_1, f_2)|^2} + 2 \sum_{\tau=1}^{N-1} \left( 1 - \frac{\tau}{N} \right) \frac{g_2(\tau; f_1)g_2(\tau; f_2)g_2(\tau; f)}{|g_3(0, 0; f_1, f_2)|^2},$$

relative to a (possibly tapered) biperiodogram.

As in the second-order case, the expressions for non-stationary processes are essentially the same as in the stationary case, with  $N$  replaced by the length of a stationary interval, and sums over all time instants with sums only over a stationary interval.

## 5. NUMERICAL EXAMPLE

### 5.1. Choice of wavelet

In the numerical simulations we have chosen a wavelet on the form (6) with a window given by a Gaussian, i.e.  $g(t) = \exp[-t^2/(2\sigma^2)]$  where  $\sigma^2$  is a user specified parameter. This choice of wavelet, the Morlet-Grossmann wavelet, is well known for its good simultaneous time and frequency resolution. The wavelet is not exactly analytic, but if we chose the parameters such that  $\eta^2\sigma^2 \gg 1$ , the non-analytic part is negligible. The parameter  $\sigma^2$  affects the time and frequency resolution. The frequency resolution is  $\Delta f/f \propto 1/(4\sigma)$  while the time resolution is  $\Delta\tau \propto \sigma/f$ . Thus the trade-off between time and frequency resolution is controlled by  $\sigma^2$ . In the following example we have used  $\eta = 4\pi$  and  $\sigma^2 = 1.5$ .

### 5.2. A piecewise stationary process

The wavelet-polyspectra give us the opportunity to analyze piecewise stationary processes, which obviously have spectral representations of the form given in (1). To demonstrate this, we provide a numerical example. The chosen signal consists of three harmonic oscillators, where the third is completely phase coupled to

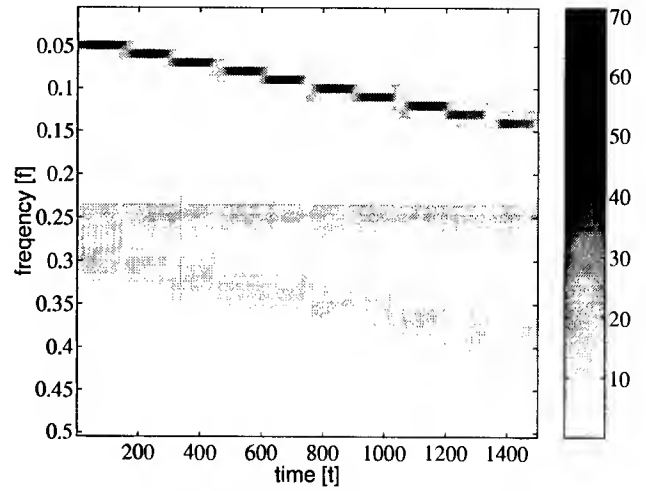


Figure 1: Magnitude squared CWT estimate  $\left| \widehat{W}_\psi(k, f) \right|^2$  of a realization  $x_n$  of the process in equation (15).

the two other. The frequency of one of the oscillators is allowed to vary with time, providing a time changing phase coupling. The signal model is

$$X_n = \sum_{i=1}^3 \cos(2\pi f_i n + \theta_i) + N_n. \quad (15)$$

Here  $f_1$  is a non-decreasing piecewise constant function,  $f_2 = 0.25$  and  $f_3 = f_1 + f_2$ . The phases  $\theta_1$  and  $\theta_2$  are independent phases drawn from a uniform distribution  $\mathcal{U}[-\pi, \pi]$  and  $\theta_3 = \theta_1 + \theta_2$ . The additive noise  $N_n$  is white, zero mean and Gaussian, with variance  $\sigma_N^2 = 0.15^2$ .

To detect phase coupling we use a wavelet based squared bicoherence estimator [5, 6]

$$\widehat{b}_w^2(f_1, f_2; L) = \frac{\left| \widehat{B}^w(f_1, f_2; L) \right|^2}{\frac{1}{2K+1} \sum_{k=L-K}^{L+K} \left| \widehat{W}_\psi(k, f_1) \widehat{W}_\psi(k, f_2) \right|^2 \widehat{S}^w(f_1 + f_2; L)}, \quad (16)$$

where we have introduced the *wavelet-bispectrum*  $B^w \triangleq M_3^w$  and the *wavelet power spectrum*  $S^w \triangleq M_2^w$ . A squared bicoherence spectrum measures the fraction of power at a given frequency due to three-wave interaction [11].

Figure 1 shows the magnitude squared CWT estimate of a realization  $x_n$  of the process in equation (15). Notice the structures corresponding to the constant fre-

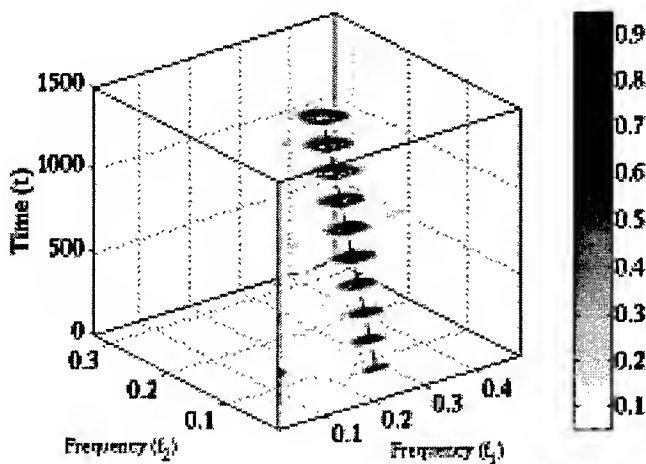


Figure 2: Estimated evolutionary bicoherence spectrum of the process in equation (15).

quency component with  $f_2 = 0.25$ , the piecewise constant frequency component  $f_1$  and the sum frequency component  $f_3 = f_1 + f_2$ . We can easily see that the period of stationarity of the process is chosen to be 150.

Figure 2 shows a contour plot of the estimate of the proposed evolutionary bicoherence spectrum, using the estimator introduced in (16). The full line passes through the true coupling frequencies. The true bicoherence value along this line is exactly 1, since all the power is due to three-wave interaction. The estimates are performed at time instants  $L$  corresponding to the midpoint of each stationary interval, and we have used  $K = 75$ . The maximum value of the estimate hits as close to the correct frequency as possible for each time instant  $L$ , and the estimated values of the maxima are about 0.97. Note that the frequency resolution of the estimate gets coarser with increasing frequency due to the constant-Q property ( $\Delta f/f = \text{const.}$ ) of the CWT.

## 6. CONCLUSION

We have proposed a definition of wavelet-polyspectra as an analysis technique for non-stationary processes. Furthermore, we have suggested wavelet-polyspectral estimators and derived important statistical properties of these for important special cases. The technique is shown to be analogous to the WOSA spectral estimation method for a fixed frequency in the second-order stationary case. The proposed estimators are shown to yield a significant variance reduction relative to the tapered periodogram and biperiodogram in the second-order and third-order case, respectively.

We have illustrated our method by a relevant numerical example. The theoretical properties of the wavelet-polyspectra and the numerical example clearly demonstrate the potential of this technique for the analysis of higher-order spectral properties of non-stationary processes.

## REFERENCES

- [1] L. Cohen, *Time-Frequency Analysis*, Prentice Hall, 1995
- [2] S. Mallat, *A Wavelet Tour of Signal Processing*, Academic Press, 1998
- [3] N. L. Gerr, Introducing a Third-Order Wigner Distribution, *Proc. IEEE*, vol. 76, pp. 290–292, March 1988
- [4] J. R. Fonollosa and C. L. Nikias, Wigner Higher Order Moment Spectra: Definition, Properties, Computation and Application to Transient Signal Analysis, *IEEE Trans. Sig. Process.*, vol. 41, pp. 245–266, January 1993
- [5] T. Dudok de Wit and V. V. Krasnosel'skikh, Wavelet Bicoherence Analysis of Strong Plasma Turbulence at the Earth's Quasiparallel Bow Shock, *Phys. Plasmas*, vol. 2, pp. 4307–4311, November 1995
- [6] B. Ph. van Milligen, C. Hidalgo and E. Sánchez, Nonlinear Phenomena and Intermittency in Plasma Turbulence, *Physical Review Letters*, vol. 74, pp. 395–398, 1995
- [7] M. B. Priestley, Evolutionary Spectra and Non-Stationary Processes, *J. Roy. Statist. Soc. Ser. B*, vol. 27, pp. 204–237, 1965
- [8] C. L. Nikias and A. P. Petropulu, *Higher-Order Spectra Analysis*, Addison-Wesley, 1993
- [9] Y. Larsen, Wavelet-Polyspectra, Master's thesis, Dept. of Physics, University of Tromsø, Norway, 1999
- [10] D. B. Percival and A. T. Walden, *Spectral Analysis for Physical Applications — Multitaper and Conventional Univariate Techniques*, Cambridge University Press, 1993
- [11] Y. C. Kim and E. J. Powers, Digital Bispectral Analysis and Its Applications to Nonlinear Wave interactions, *IEEE Trans. Plasma Science*, vol. 2, pp. 120–131, June 1979

# ADAPTIVE SEISMIC COMPRESSION BY WAVELET SHRINKAGE

M.F. Khène, S.H. Abdul-Jauwad

King Fahd University of Petroleum & Minerals  
Electrical Engineering Department  
Dhahran 31261, Saudi Arabia  
{samara, mfkhene@kfupm.edu.sa}

## ABSTRACT

In this paper, a sophisticated adaptive seismic compression method is presented based on *wavelet shrinkage*. Our approach combines a time-scale transform with an adaptive non-linear statistical method. First, a discrete 2-D biorthogonal Discrete Wavelet Transform (DWT) is applied to the multi-channel seismic signals to generate a sparse multiresolution (subband) decomposition. Compression is then achieved by shrinking the detail wavelet coefficients using a scale-dependent non-linear soft-thresholding rule. The adaptive scale-dependent thresholds are determined by minimizing the *Stein's Unbiased Risk Estimate (SURE)*. The proposed compression procedure is tested on marine seismic data from the *Midyan* basin (*Red Sea, Saudi Arabia*).

## 1. INTRODUCTION

Seismic compression is a key technology for managing seismic data in a world of ever increasing data volumes to maintain productivity without compromising interpretation results. By storing data in a format that requires less space than the original data volume, seismic compression provides greater flexibility in managing local or remote server disk space as well as reducing network traffic. Seismic compression not only enables explorationists to maximize the value of the information technology infrastructure, but it encourages innovative interpretation workflow to leverage the vast information content in massive seismic data. Compression thus helps in maintaining or exceeding current productivity levels. Recently, seismic compression has benefited from the advent of wavelets [1], which offer mathematical constructions with a great potential in statistical methodology [2]. Wavelet transforms have been applied extensively in diverse applications including data compression and denoising, image analysis, economics and statistics [3].

In this paper, a sophisticated wavelet-based compression technique is proposed. Both the transform and the compression stages are matched up in view to improving the overall performance. The rationale of our approach is first to generate a (near)-sparse representation of the data in the wavelet domain then to threshold an important part of the coefficients without losing substantial information. In order to exploit the multi-channel seismic data correlation in space and time directions, a 2-D biorthogonal DWT is used. For seismic interpretation, the visual aspect of seismic signals is of utmost importance. This factor is taken into account by judiciously selecting both the wavelets and the thresholding procedure. Thus, a DWT using

long wavelet filters from the *Cohen-Daubechies-Feauveau* (CDF) class [4] is intimately associated with a non-linear smooth operator, namely wavelet *shrinkage*. Biorthogonal wavelets offer a good trade-off between the support size, the number of vanishing moments and regularity. In other words, the DWT is computed efficiently while preventing the appearance of artifacts in the reconstructed data. In addition, the sparsity of the multiresolution decomposition is best exploited by wavelet shrinkage. The latter consists of applying a soft-thresholding rule to all the wavelet coefficients but those belonging to the lowest resolution subband. Indeed, the latter is merely a smooth scaled version of the input data and carries the essence of the data. Moreover, it has coefficients of much smaller magnitude than those of the detail subbands do. Thus, this makes its contribution to the compression gain marginal. The values of the scale-dependent thresholds are determined by minimizing the SURE [5][6]. The proposed compression procedure is tested on marine seismic data from the *Midyan* basin (*Red Sea, Saudi Arabia*) [7].

## 2. MULTICHANNEL SEISMIC SIGNALS

Oil and gas are usually buried deep within the earth, often miles below the surface. Most of the easy or shallow oil has been found. A number of exploration methods are available but in general only the modern seismic reflection method comes close to providing both the ability to see down to great depths and to see the details of the subsurface needed to locate many hydrocarbons [8]. Seismic data stem from a multiscale non-linear distributed parameters remote system, i.e. the earth. In a typical scenario, a spatially distributed acoustical signal is generated by a source (e.g. dynamite) located at the surface of the earth. The generated waves propagate downward, undergo reflection at contacts with different acoustic impedance, and are recorded by an array of seismometers at the surface. This provides the multi-channels discrete seismic signals that are mapped into representations of the earth's interior properties. The underlying complex process is referred to as seismic imaging. The latter is intended to find earth models that explain (or best fit) seismic observations. Seismic signals are commonly displayed using a variable-area and/or a variable-density mode [8].

## 3. MULTIREOLUTION DECOMPOSITION

### 3.1 2-D Wavelet Bases

There are two different ways to build a wavelet basis for a 2-D space, say  $t$  and  $x$ . The standard dyadic construction consists of

all possible tensor products of 1-D wavelet and scaling basis functions defined respectively as:

$$\psi_{jk}(t) = 2^{\frac{j}{2}} \psi_{jk}(2^j t - k) \quad (1a)$$

$$\phi_{jk}(t) = 2^{\frac{j}{2}} \phi_{jk}(2^j t - k) \quad (1b)$$

However, despite its simplicity, the construction that requires different scale indices for each direction, does not benefit from the recursive *Mallat* algorithm [9]. Indeed, for an  $m \times m$  matrix data the standard dyadic decomposition requires  $4(m^2 - m)$  assignment operations against  $8/3(m^2 - 1)$  only for the nonstandard one [10]. Consequently, in the sequel the nonstandard dyadic 2-D decomposition is adopted. It consists of defining a 2-D scaling function, using a unique scale index  $j$  as:

$$\phi_{jkk'}(t, x) = \phi_{jk}(t) \phi_{jk'}(x) \quad (2)$$

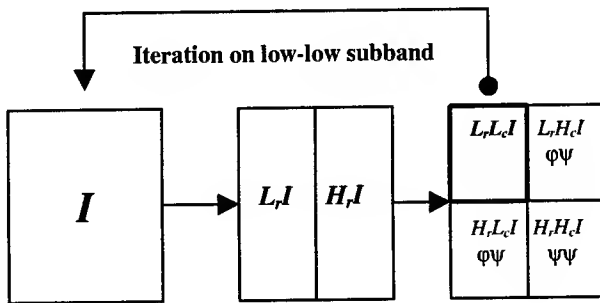
and three 2-D wavelet functions at each scale given by:

$$\psi_{jkk'}^V(t, x) = 2^{\frac{j}{2}} \psi_{jk}(2^j t - k, 2^j x - k') \quad (3a)$$

$$\psi_{jkk'}^H(t, x) = 2^{\frac{j}{2}} \psi_{jk}(2^j t - k, 2^j x - k') \quad (3b)$$

$$\psi_{jkk'}^D(t, x) = 2^{\frac{j}{2}} \psi_{jk}(2^j t - k, 2^j x - k') \quad (3c)$$

These three *anisotropic* wavelets extract matrix data details at different scales and orientations, whereas the scaling function yields a smoothed low-resolution version of the input data. Indeed, starting at scale  $j$ , the multiresolution decomposition yields four double-scaled half-resolution panels at scale  $j-1$ . One of them represents a smoothed version of the data while the remaining ones contain detail wavelet coefficients corresponding to the  $\{\psi^V, \psi^H, \psi^D\}$  wavelet functions that are respectively oriented vertically, horizontally and diagonally. The result of the nonstandard dyadic 2-D DWT is usually displayed in four panels as in Fig.1.



**Figure 1.** Nonstandard dyadic 2-D wavelet multiresolution decomposition.  $L$  and  $H$  stand for low- and high-pass wavelet filters, and the subscripts  $r$  and  $c$  stand for row and column respectively

### 3.2 Biorthogonal Wavelet Bases

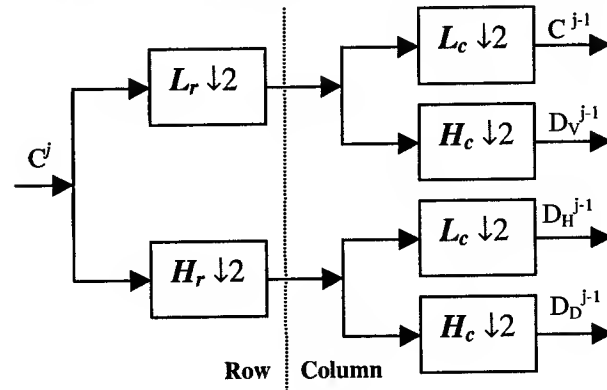
There are three main categories of wavelet bases, namely the orthogonal, the semi-orthogonal and the biorthogonal. Limiting ourselves to orthogonal wavelet bases can be overly restrictive because except for the *Haar* basis, there are no other bases, which are compactly supported and symmetric. The relaxation of orthogonality constraint has many benefits that improve the performance of the wavelet transform while still being implemented with the *Mallat* algorithm. In particular, biorthogonal wavelets offer a good trade-off between the support size, the number of vanishing moments and regularity. In term of digital filters, the biorthogonal transform uses different Finite Impulse Response (FIR) wavelet filters in the decomposition and reconstruction stages. This provides more flexibility in the design of the transform and its inverse [11]. Moreover, FIR filters are preferred because they guarantee a linear phase, which is a very desirable property that prevents from the appearance of artifacts in the reconstructed data. Therefore, the biorthogonal transform uses dual wavelet and dual scaling functions related to the primal ones by:

$$\left\{ \begin{array}{l} \langle \phi_{jk} | \tilde{\psi}_{jk'} \rangle = 0 \\ \langle \psi_{jk} | \tilde{\phi}_{jk'} \rangle = 0 \end{array} \right\} \text{ for all } j, k, k' \quad (4)$$

In this contribution long biorthogonal wavelets filters of the *CDF* class are used [4].

### 3.3 Nonstandard Dyadic 2-D Decomposition

The extension to 2-D separable biorthogonal bases is straightforward. Indeed, by alternating the 1-D wavelet filtering operations on rows and columns a 2-D dyadic nonstandard decomposition is generated. This scheme is implemented with a two-channel filter bank [11], where the low-pass and high-pass filters represent the scaling and the wavelet functions respectively. First, a one step low pass ( $L$ ) and high pass ( $H$ ) filtering is performed on each row of the matrix  $I$ . Next, the same filters are applied to each column of the resulting matrix. The whole process is applied recursively to the quadrant containing averages in both directions, i.e.  $L_r L_c I$  panel. The resulting recursive decomposition is illustrated by Fig.1. and the implementation of the one-stage 2-D DWT is depicted by Fig.2.



**Figure 2.** One-stage non-standard 2-D DWT

## 4. WAVELET SHRINKAGE

### 4.1 Motivation

Wavelet compression is best understood from an approximation viewpoint. In a wavelet decomposition, each wavelet picks up information about the data at a given location  $k$  and at a given resolution or scale  $j$ . Thus, the wavelet transform allows us to focus on the most relevant part of the data provided the wavelet bases fit the input data. Consequently, the resulting wavelet coefficients drop off rapidly yielding a (near)-sparse data representation. This is also known as energy compaction property. Wavelet thresholding constitutes thus a natural choice to perform compression. It is a simple yet a very efficient procedure for keeping the most important coefficients that will be used in reconstruction. An intuitive way to achieving thresholding consists of applying a *keep-or-kill* rule referred to as *hard-thresholding*. However, a *soft-thresholding* is preferred because of various advantages. From a visual point of view, the reconstructed data offer a more pleasant aspect, and do not exhibit visible artifacts. This is crucial in the case of seismic data interpretation. From a statistical point of view, soft-thresholding uses a continuous function, leading to simple data driven selection of the thresholds. In fact, the selection of the thresholds is a very delicate and important statistical problem. On one hand, killing too many wavelet coefficients may lead to an important bias in the reconstructed data. On the other hand, small thresholds lead to a poor compression gain. Thus, threshold selection should strike the balance between closeness to fit between the original and the reconstructed data and the degree of sparsity of the wavelet coefficients. We propose to achieve compression through an adaptive soft-thresholding procedure, referred to as wavelet shrinkage. The selection of the optimal threshold for all the scales but the coarsest one is accomplished by minimizing the SURE. The resulting nonlinear thresholding operator is called *SureShrink* [5].

### 4.2 SURE Principle

SURE has been initiated by *Stein* for mean estimation of a multivariate normal distribution [12] and has been successfully applied for function smoothing by *Donoho* [5]. The foundation of the SURE principle is based on the fact that for nearly arbitrary nonlinear biased estimator, the loss or risk can be estimated unbiasedly. In the sequel, we outline the SURE principle for the general case then in the next paragraph we show how to derive the *SureShrink* operator.

Consider an empirical data vector  $y$  of dimension  $N$  given by

$$y_i = f_i + e_i, \quad i = 0, 1, \dots, N-1 \quad (5)$$

where  $f_i$  are samples of the deterministic function  $f$  and  $e$  is Gaussian white noise with independent identical distribution (i.i.d)  $N(0, \sigma)$ .

The objective is to find the best estimate of the function  $f$  in the mean square sense by minimizing the *Mean Square Error* (MSE) risk defined as,

$$R(\hat{f}, f) = \frac{1}{N} \|\hat{f} - f\|^2 = \frac{1}{N} \sum_{i=0}^{N-1} (\hat{f}_i - f_i)^2 \quad (6)$$

However, the main drawback of the MSE risk is that in practice, it can never be computed exactly because it relies on the unknown exact value of the function  $f$ . Thus in practical situations this MSE has to be estimated. The SURE principle stipulates that if we consider the following estimate for the unknown function  $f$ ,

$$\hat{f}(y) = y + g(y) \quad (7)$$

where  $g(y)$  is a weakly differentiable function from  $R^N$  to  $R^N$ , then an unbiased estimator for the MSE risk is the SURE defined as [12]:

$$R^{SURE}(\hat{f}(y), f) = N + \|\hat{g}(y)\|^2 + 2\nabla_y \cdot \hat{g}(y) \quad (8)$$

where  $\nabla$  is the vector differential operator of first partial derivatives, i.e.,

$$\nabla_y \cdot \hat{g}(y) = \sum_{i=0}^{N-1} \frac{\partial}{\partial y_i} g_i \quad (9)$$

### 4.3 Adaptive Wavelet Shrinkage with SureShrink

There are two main classes of wavelet shrinkage regarding whether the threshold is single and global or scale-dependent and adaptive. Our approach consists of deriving a scale-dependent threshold  $\lambda_j$  according to the following soft-thresholding rule:

$$g_\lambda(d) = \text{sgn}(d_k^j) (|d_k^j| - \lambda^j) I(|d_k^j| > \lambda^j) \quad (10)$$

For a given detail subband at resolution  $j$ , the shrinkage operator  $g_\lambda(d)$  kills all those coefficients below the threshold  $\lambda^j$  and pulls towards the origin the surviving ones by an amount equals to the threshold. The different scale-dependent thresholds stem from the minimization of the SURE, i.e.,

$$\lambda^j = \arg \min_{\lambda \geq 0} R_\lambda^{SURE}(\lambda^j, d_k^j) \quad (11)$$

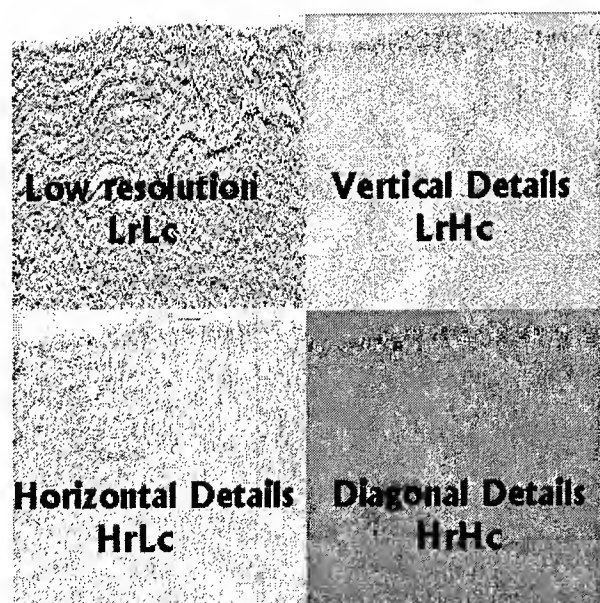
where the SURE for a soft-threshold estimator is given by [5],

$$R_\lambda^{SURE} = 2^j - 2I\{|d_k^j| \leq \lambda^j\} + \sum_{i=0}^{2^j-1} \min\{|d_k^j|, \lambda^j\}^2 \quad (12)$$

Note that the underlying optimization problem is straightforward and the computational effort is of order  $2^j \log(2^j)$  as a function of the subband size  $2^j$  [5].

## 5. EXPERIMENTAL RESULTS

A migrated marine seismic profile from the *Midyan* basin in the *Red Sea* is used to demonstrate our adaptive seismic compression by wavelet shrinkage. The discrete 2-D seismic data consist of a collection of 2838 seismic signals (traces) of 2.5 seconds length each sampled at 4 milliseconds. The traces correspond to the *Common MidPoints* (CMP), i.e., successive reflection points midway between the different seismic source locations and the seismometers. The data can be regarded as a (626x2838) matrix of floats entries. The variable density display mode is used to represent the profile, which can be thought of as a transversal section of the prospected area along the seismic line. First we have applied a 2-D DWT with asymmetric long biorthogonal wavelet filters  $CDF(6,8)$  where the numbers of vanishing moments for the synthesis and analysis wavelets are 6 and 8 respectively. The reader may be wondering why the reconstruction wavelet filters are shorter and have less vanishing moments than the decomposition ones. First note that this is made possible thanks to the flexibility of the biorthogonal wavelet transform. Second, the objectives of the decomposition and reconstruction stages differ. Indeed, the main concern of the wavelet decomposition is to pack the energy of the input data in fewer wavelet coefficients. The higher the number of vanishing moments is, the better the energy compaction would be. From the reconstruction side, we wish to use smooth wavelets to mask the errors introduced by the wavelet shrinkage and to get less annoying visual artifacts. Furthermore, the reconstruction time should be shorter than the decomposition one because, in practice, the data set is compressed once but may be decompressed several times. A three-level multiresolution decomposition has been performed yielding nine detail subbands and one low-resolution subband. The four subbands of the first level are displayed in Fig.3.



**Figure 3.** First level nonstandard dyadic 2-D DWT multiresolution decomposition of the *Midyan* section

Next, we have applied the *SureShrink* operator to the nine details subbands. The experimental results are displayed in Fig.4. Though, almost 82% of the wavelet coefficients have been killed by shrinkage, 95% of the data energy is recovered in the reconstructed data. Furthermore, the difference section exhibits random noise. Thus, an appreciable filtering effect has also been produced by the compression.

## 6. CONCLUSIONS

In this work, a sophisticated adaptive seismic compression technique was presented. A time-scale transform was associated with a non-linear statistical method. A pair of different asymmetric biorthogonal wavelet filters was selected to achieve different targets for the compression and decompression processes. Analysis wavelets with more vanishing moments were used to ensure a maximum energy compaction of the input data. This made compression by thresholding a very natural and efficient means. Based on *SURE*, the scale-dependent thresholds were determined and then the *SureShrink* operator was applied to each detail subband to kill insignificant coefficients. The experimental results show that the proposed approach does not introduces visible artifacts while achieving a relatively high compression gain.

## 7. ACKNOWLEDGMENTS

The authors wish to gratefully acknowledge *King Fahd University of Petroleum & Minerals* for its support and *Saudi ARAMCO* for providing the data.

## 8. REFERENCES

- [1] Bosman C. and Reiter, E. "Seismic data compression Using wavelet transform." *63<sup>rd</sup> SEG Expanded Abstracts*, 1261-1264, 1993
- [2] Vidakovic B. *Statistical modeling by wavelets*. Wiley Series in Probability and Statistics. John Wiley & Sons, 1999.
- [3] Meyer Y. *Wavelets: Algorithms and applications*. SIAM 1993.
- [4] Cohen A., Daubechies I., Feauveau J. "Biorthogonal bases of compactly supported wavelets." *Commun. on Pure Appl. Math.*, 45:485-560, 1992.
- [5] Donoho D. and Johnstone I. "Adapting to unknown smoothness by wavelet shrinkage." *Journ. of the Amer. Statistical assoc.*, 90:1200-1224, 1995.
- [6] Misitti M., Missiti Y., Oppenheim. G and Poggi J-M. *Matlab wavelet toolbox*. The MathWorks, Inc, 1997.
- [7] Mougnot D. and Al-Shakhis A. "Depth imaging a pre-salt faulted block: A case study from the *Midyan* basin (*Red Sea*)." *Saudi Aramco Jour. of Tech.* Fall 1998.
- [8] Sheriff R. *Geophysical methods*. Prentice Hall, 1989.
- [9] Mallat S. *A Wavelet tour of signal processing*. Academic Press, 1998.
- [10] Stollinz E., DeRose T. and Salesin D. *Wavelets for computer graphics: Theory and applications*. Morgan Kaufmann Publishers, Inc., San Francisco, 1996.
- [11] Vetterli M. and Kovacevic J. *wavelets and subband coding*. Prentice hall PTR, New Jersey, 1995.
- [12] Stein C. "Estimation of the mean of a multivariate normal distribution" *The Annals of Statistics*, 9(6): 1135-1151, 1981.



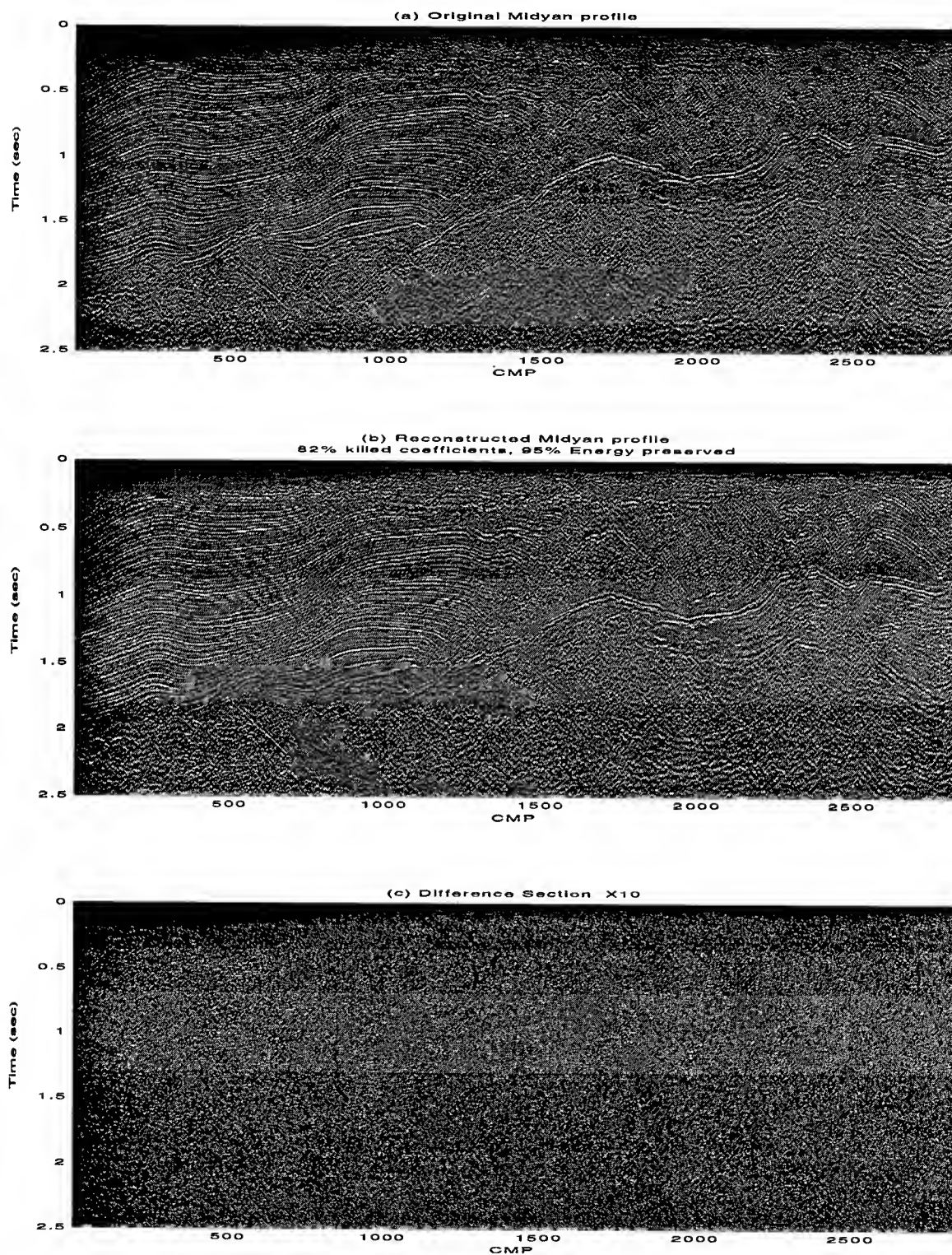


Figure 4. Experimental results



# REPRESENTATIONS OF STOCHASTIC PROCESSES USING COIFLET-TYPE WAVELETS

*Dong Wei and Haiguang Cheng*

Center for Telecommunications and Information Networking  
Department of Electrical and Computer Engineering  
Drexel University  
Philadelphia, PA 19104 U.S.A.

E-mails: wei@ece.drexel.edu, hgcheng@io.ece.drexel.edu

## ABSTRACT

The wavelet series expansion requires a high computational complexity in computing the scaling coefficients at the finest scale by means of projection in order to realize the Mallat algorithm to compute the wavelet coefficients at coarser scales. We propose a fast and practical algorithm to approximate the wavelet series expansion. The algorithm is based on sampling and reconstruction with coiflet-type wavelets, which possess vanishing moments on both scaling function and wavelet. We evaluate the performance of the algorithm by establishing the convergence rates and asymptotic forms for the mean-square errors in the scaling coefficients and wavelet coefficients of the synthesized stochastic process.

## 1. INTRODUCTION

During the past decade, the theory of wavelets has established itself firmly as one of the most successful mathematical tools for a broad range of signal processing applications, such as image data compression, noise reduction, and singularity detection. A fundamental and important problem in wavelet-based multiresolution approximation theory is to measure the decay of the approximation error as resolution increases. The convergence properties and rates for the wavelet series expansion (WSE) of stochastic processes have been studied in [1]. The WSE requires the computation of the scaling coefficients at the finest scale by means of projection in order to realize the Mallat algorithm to compute the wavelet coefficients at coarser scales. Since, in practice, uniform samples of signals rather than their analytic forms are often available, the projection-based implementation of the WSE requires

numerical integrals to approximate the scaling coefficients, which are computationally expensive. Therefore, such an implementation is far from practical.

In this paper, we propose a fast and practical algorithm to accurately approximate the WSE. The algorithm is based on sampling and reconstruction with coiflet-type wavelets, which possess vanishing moments on both scaling function and wavelet. We study the resulting wavelet representations of stochastic processes and evaluate the performance of the algorithm by establishing the convergence rates and asymptotic forms for the mean-square errors in the scaling coefficients and wavelet coefficients of the synthesized stochastic process. This work is parallel to the result in [1] and can be viewed as a generalization of the result on approximated WSE for deterministic functions [2].

The following simplified notation is used in the paper:

$$\int \equiv \int_{-\infty}^{\infty}, \quad \sum \equiv \sum_{n=-\infty}^{\infty}.$$

Due to space limitation, the proofs of some of the presented results are not given in this paper.

## 2. BACKGROUND

### 2.1. Wavelet Representations

First, we briefly review wavelet-based multiscale representations. A continuous-time, real-valued stochastic process  $X(t)$  can be approximated by its finite-scale wavelet series expansion:

$$\begin{aligned} \hat{X}_{i_1}(t) &= \sum_k s_{i_1}[k] \phi_{i_1,k}(t) \\ &= \sum_k s_{i_0}[k] \phi_{i_0,k}(t) + \sum_{i=i_0}^{i_1-1} \sum_k w_i[k] \psi_{i,k}(t) \quad (1) \end{aligned}$$

This work was supported by Defense Advanced Research Project Agency under grant F30602-00-2-0501.

where we have used the short-hand notation

$$\psi_{i,k}(t) = 2^{i/2} \psi(2^i t - k)$$

for the dilated and translated versions of  $\psi$ , and applied the notation to  $\phi_{i_1,k}(t)$  and  $\phi_{i_0,k}(t)$  similarly. The functions  $\phi$  and  $\psi$  are the scaling function and the wavelet, respectively. In this paper, we only consider two-channel, compactly supported, orthonormal scaling functions and wavelets [3], [4]. The scaling coefficients at scale  $2^i$  are defined as

$$s_i[k] = \int X(t) \phi_{i,k}(t) dt$$

and the wavelet coefficients at scale  $2^i$  are defined as

$$w_i[k] = \int X(t) \psi_{i,k}(t) dt.$$

Given  $\{s_{i_1}[k] : k \in \mathbf{Z}\}$ , the set of scaling coefficients at the finest scale  $2^{i_1}$ , the wavelet coefficients  $\{w_i[k] : k \in \mathbf{Z}\}$  can be computed efficiently in a recursive fashion via the Mallat algorithm [5]:

$$s_i[k] = \sum_n h[2k - n] s_{i+1}[n]$$

$$w_i[k] = \sum_n g[2k - n] s_{i+1}[n]$$

for  $i = i_1 - 1, i_1 - 2, \dots, i_0$ , where  $h[n]$  and  $g[n]$  are a pair of finite-impulse-response conjugate quadrature filters [4].

## 2.2. Coiflet-Type Wavelets

A *coiflet-type wavelet system* of order  $L$  satisfies the Coifman criterion; i.e., the first  $L$  wavelet moments vanish

$$m_\psi[l] = \int t^l \psi(t) dt = 0 \quad \text{for } l = 0, 1, \dots, L - 1$$

and the second to the  $L$ th scaling function moments vanish

$$m_\phi[l] = \int t^l \phi(t) dt = \delta[l] \quad \text{for } l = 0, 1, \dots, L - 1$$

where  $\delta[\cdot]$  denotes the Kronecker delta sequence. Examples of orthonormal coiflet-type wavelets include the original coiflets [6] and their generalized versions [7]. If a coiflet-type wavelet system is used in the WSE of a deterministic signal  $X(t)$ , then the uniform signal samples approximate the scaling coefficients accurately at a sufficiently fine scale [4]:

$$s_i[k] = 2^{-i/2} X(2^{-i}k) + \mathcal{O}(2^{-iL}).$$

Such a property is highly appealing in digital signal processing applications, where samples of signals are processed digitally.

## 3. A FAST AND PRACTICAL ALGORITHM FOR WAVELET REPRESENTATIONS

The projection-based approximation given in (1) has a limitation in reality. In many practical applications, the initial set of scaling coefficients  $\{s_{i_1}[k] : k \in \mathbf{Z}\}$  is difficult and computationally costly, if not impossible, to obtain. In most cases, uniform samples of signals rather than analytic function forms are available.

To overcome this challenging difficulty, we propose a fast algorithm for computing the approximated scaling and wavelet coefficients in the WSE based on an  $L$ th-order coiflet-type wavelet system:

- (i) the approximated scaling coefficients at scale  $2^i$  are the uniform samples of  $X(t)$  with a sampling period  $2^{-i}$ :

$$\hat{s}_i[k] = 2^{-i/2} X(2^{-i}k)$$

- (ii) the approximated wavelet coefficients at scale  $2^i$  are computed from the approximated scaling coefficients at scale  $2^{i+1}$  as

$$\hat{w}_i[k] = \sum_n g[2k - n] \hat{s}_{i+1}[n].$$

From the above algorithm, we know that the approximated wavelet coefficients at scale  $2^i$  are the filtered and decimated versions of the uniform samples of the stochastic process with a sampling period  $2^{-i-1}$ .

An alternative to the above algorithm would be to use the densest samples  $\{2^{-i_1/2} X(2^{-i_1}k) : k \in \mathbf{Z}\}$  as an initial set of scaling coefficients to trigger the Mallat algorithm to compute wavelet coefficients at all coarser scales. However, the proposed algorithm possesses the following two advantages over the alternative:

- since the proposed algorithm is nonrecursive, the approximation error at the finest scale  $2^{i_1}$  does not propagate across the coarser scales;
- since no computation is required for scaling coefficients, the computational complexity of the proposed algorithm is half of the computational complexity of the Mallat algorithm.

Let  $R_{XX}(s, t) = E\{X(s)X(t)\}$  denote the autocorrelation function of the stochastic process  $X(t)$ . We assume that  $R_{XX}$  is sufficiently smooth in the sense that

$$R_{XX}^{(m,n)}(s, t) = \frac{\partial^{m+n} R_{XX}(s, t)}{\partial s^m \partial t^n}$$

exists and is finite for any  $s, t \in \mathbf{R}$  and any  $m, n \in \mathbf{Z}$ ,  $m + n \leq K$ , where  $K$  is a sufficiently large integer.

The following two propositions evaluate the approximation accuracy of the proposed algorithm in terms of mean-square error (MSE).

**Proposition 1.** The MSE in the scaling coefficient  $\hat{s}_i[k]$  has the asymptotic form

$$\begin{aligned} & E\{(s_i[k] - \hat{s}_i[k])^2\} \\ &= 2^{-i(2L+1)} \binom{2L}{L} \frac{(m_\phi[L])^2}{L!} R_{XX}^{(L,L)}\left(\frac{k}{2^i}, \frac{k}{2^i}\right) \\ & \quad + \mathcal{O}(2^{-i(2L+2)}). \end{aligned}$$

**Proof.** It follows that

$$\begin{aligned} & E\{(s_i[k] - \hat{s}_i[k])^2\} \\ &= E\{(s_i[k])^2\} - 2E\{s_i[k]\hat{s}_i[k]\} + E\{(\hat{s}_i[k])^2\} \\ &= E\left\{\int \int X(t_1)X(t_2)\phi_{i,k}(t_1)\phi_{i,k}(t_2) dt_1 dt_2\right\} \\ & \quad - 2E\left\{2^{-i/2}X(2^{-i}k) \int X(t)\phi_{i,k}(t) dt\right\} \\ & \quad + E\{2^{-i}X(2^{-i}k)X(2^{-i}k)\} \\ &= 2^i \int \int R_{XX}(t_1, t_2)\phi(2^i t_1 - k)\phi(2^i t_2 - k) dt_1 dt_2 \\ & \quad - 2 \int R_{XX}(t, 2^{-i}k)\phi(2^i t - k) dt \\ & \quad + 2^{-i}R_{XX}(2^{-i}k, 2^{-i}k) \\ &= 2^{-i} \int \int R_{XX}\left(\frac{s_1+k}{2^i}, \frac{s_2+k}{2^i}\right) \phi(s_1)\phi(s_2) ds_1 ds_2 \\ & \quad - 2^{-i+1} \int R_{XX}\left(\frac{s+k}{2^i}, \frac{k}{2^i}\right) \phi(s) ds \\ & \quad + 2^{-i}R_{XX}(2^{-i}k, 2^{-i}k). \end{aligned}$$

By taking a Taylor series expansion of  $R_{XX}$  at  $(2^{-i}k, 2^{-i}k)$ , we have

$$\begin{aligned} & R_{XX}\left(\frac{s_1+k}{2^i}, \frac{s_2+k}{2^i}\right) \\ &= \sum_{l=0}^{2L} \sum_{n=0}^l \binom{l}{n} \frac{R_{XX}^{(n,l-n)}(2^{-i}k, 2^{-i}k)}{2^{il}l!} s_1^n s_2^{l-n} \\ & \quad + \mathcal{O}(2^{-i(2L+1)}) \end{aligned}$$

and

$$\begin{aligned} R_{XX}\left(\frac{s+k}{2^i}, \frac{k}{2^i}\right) &= \sum_{l=0}^{2L} \frac{R_{XX}^{(l,0)}(2^{-i}k, 2^{-i}k)}{2^{il}l!} s^l \\ & \quad + \mathcal{O}(2^{-i(2L+1)}). \end{aligned}$$

Using the vanishing moment property of coiflet-type wavelets, we infer that

$$E\{(s_i[k] - \hat{s}_i[k])^2\}$$

$$\begin{aligned} &= \sum_{l=0}^{2L} \sum_{n=0}^l \binom{l}{n} \frac{R_{XX}^{(n,l-n)}(2^{-i}k, 2^{-i}k)}{2^{i(l+1)}l!} m_\phi[n]m_\phi[l-n] \\ & \quad - 2 \sum_{l=0}^{2L} \frac{R_{XX}^{(l,0)}(2^{-i}k, 2^{-i}k)}{2^{i(l+1)}l!} m_\phi[l] \\ & \quad + 2^{-i}R_{XX}(2^{-i}k, 2^{-i}k) + \mathcal{O}(2^{-i(2L+2)}) \\ &= \sum_{l=L}^{2L} \frac{R_{XX}^{(0,l)}(2^{-i}k, 2^{-i}k) + R_{XX}^{(l,0)}(2^{-i}k, 2^{-i}k)}{2^{i(l+1)}l!} m_\phi[l] \\ & \quad + \binom{2L}{L} \frac{R_{XX}^{(L,L)}(2^{-i}k, 2^{-i}k)}{2^{i(2L+1)}(2L)!} (m_\phi[L])^2 \\ & \quad - 2 \sum_{l=L}^{2L} \frac{R_{XX}^{(l,0)}(2^{-i}k, 2^{-i}k)}{2^{i(l+1)}l!} m_\phi[l] + \mathcal{O}(2^{-i(2L+2)}) \\ &= \binom{2L}{L} \frac{R_{XX}^{(L,L)}(2^{-i}k, 2^{-i}k)}{2^{i(2L+1)}(2L)!} (m_\phi[L])^2 \\ & \quad + \mathcal{O}(2^{-i(2L+2)}). \end{aligned}$$

□

**Proposition 2.** The MSE in the wavelet coefficient  $\hat{w}_i[k]$  has the asymptotic form

$$\begin{aligned} E\{(w_i[k] - \hat{w}_i[k])^2\} &= C_\psi \cdot 2^{-i(4L+1)} \cdot R_{XX}^{(2L,2L)}(0,0) \\ & \quad + \mathcal{O}(2^{-i(4L+2)}) \end{aligned}$$

where

$$C_\psi = \binom{4L}{2L} \binom{2L}{L}^2 \frac{(m_\psi[L])^2}{2^{2L}(4L)!}.$$

Proposition 2 can be proved in a similar way to Proposition 1.

The above two propositions indicate that the MSE in the wavelet coefficients decays faster than the MSE in the scaling coefficients as the product  $iL$  increases.

#### 4. WAVELET REPRESENTATIONS OF STOCHASTIC PROCESSES

We study the approximation accuracy of the wavelet representations based on the proposed algorithm.

The sequence of stochastic processes

$$X_i(t) = \sum_k \hat{s}_i[k] \phi_{i,k}(t)$$

for  $t \in [a, b]$  and  $i \in \mathbb{Z}$ , can be viewed as successive approximations of  $X(t)$  over the interval  $[a, b]$  using the dilated and translated scaling functions of some  $L$ th-order coiflet-type wavelet system as the interpolants. The following proposition expresses the auto-correlation function of the process  $X_i(t)$  and the cross-correlation function of  $X_i(t)$  and  $X(t)$  asymptotically in terms of the auto-correlation function of  $X(t)$ .

**Proposition 3.** For any  $s, t \in [a, b]$ ,

$$\begin{aligned} & R_{X_i X_i}(s, t) \\ &= R_{XX}(s, t) \\ &+ \frac{R_{XX}^{(L,0)}(s, t) \rho_L(2^i s) + R_{XX}^{(0,L)}(s, t) \rho_L(2^i t)}{(-1)^L 2^{iL} L!} \\ &+ \mathcal{O}(2^{-i(L+1)}) \end{aligned} \quad (2)$$

and

$$\begin{aligned} R_{XX_i}(s, t) &= R_{XX}(s, t) + \sum_{l=L}^{2L} \frac{R_{XX}^{(0,l)}(s, t) \rho_l(2^i t)}{(-1)^l 2^{il} l!} \\ &+ \mathcal{O}(2^{-i(L+1)}) \end{aligned} \quad (3)$$

where  $\rho_l$  is a periodic function with unit period and is given by

$$\rho_l(t) = \sum_k (t - k)^l \phi(t - k).$$

We use the integrated MSE

$$e_i^2 \triangleq E \left\{ \int_a^b [X(t) - X_i(t)]^2 dt \right\} \quad (4)$$

to measure the approximation accuracy at scale  $2^i$  and evaluate the asymptotic approximation performance in the next proposition.

**Proposition 4.** The integrated MSE possesses the asymptotic form

$$e_i^2 = C_\phi \cdot 2^{-2iL} \cdot \int_a^b R_{XX}^{(L,L)}(t, t) dt + \mathcal{O}(2^{-i(2L+1)})$$

where the constant  $C_\phi$  is given by

$$C_\phi = \frac{1}{(L!)^2} \int_0^1 \rho_L^2(t) dt. \quad (5)$$

**Proof.** The MSE can be rewritten as

$$e_i^2 = \int_a^b [R_{XX}(t, t) - 2R_{XX_i}(t, t) + R_{X_i X_i}(t, t)] dt. \quad (6)$$

For  $i$  sufficiently large, the  $\mathcal{O}(2^{-i(2L+1)})$  error terms in (2) and in (3) become negligible, and we can use the pointwise estimates in Proposition 3 to obtain the asymptotic form of the error in (6),

$$\begin{aligned} & \lim_{i \rightarrow \infty} \frac{e_i^2}{2^{-2iL}} \\ &= \lim_{i \rightarrow \infty} \frac{1}{(L!)^2} \int_a^b R_{XX}^{(L,L)}(t, t) \rho_L^2(2^i t) dt \\ &= \lim_{i \rightarrow \infty} \frac{1}{(L!)^2} \sum_{m=2^i a+1}^{2^i b} \int_{\frac{m-1}{2^i}}^{\frac{m}{2^i}} R_{XX}^{(L,L)}\left(\frac{m}{2^i}, \frac{m}{2^i}\right) \rho_L^2(2^i t) dt \end{aligned}$$

$$\begin{aligned} &= \lim_{i \rightarrow \infty} \frac{1}{(L!)^2} \left[ \sum_{m=2^i a+1}^{2^i b} R_{XX}^{(L,L)}\left(\frac{m}{2^i}, \frac{m}{2^i}\right) \right] \\ &\quad \times \left[ \int_0^{2^{-i}} \rho_L^2(2^i t) dt \right] \\ &= \frac{1}{(L!)^2} \cdot \left[ \int_a^b R_{XX}^{(L,L)}(t, t) dt \right] \cdot \left[ \int_0^1 \rho_L^2(t) dt \right]. \end{aligned}$$

□

If  $R_{XX}^{(L,L)}(t, t)$  decays sufficiently fast towards infinities in the sense that

$$\left| \int R_{XX}^{(L,L)}(t, t) dt \right| < \infty,$$

then Proposition 4 holds for the limiting cases  $a = -\infty$  and/or  $b = \infty$ .

Since a deterministic function  $X(t)$  may be considered as a degenerate stochastic process with an auto-correlation function  $R_{XX}(s, t) = X(s)X(t)$ , Proposition 4 can be viewed as a generalization of the result on sampling-based approximation of deterministic functions in [2], which corresponds to the special case  $R_{XX}^{(L,L)}(t, t) = [X^{(L)}(t)]^2$ . Indeed, the above constant  $C_\phi$  is identical to the asymptotic constant in the deterministic case [2]. Therefore,  $C_\phi$  can be computed via the algorithm described in [2], provided that the filter  $h$  is known.

## 5. WAVELET REPRESENTATIONS OF STATIONARY STOCHASTIC PROCESSES

Let  $R_{XX}(\tau) = E\{X(t-\tau)X(t)\}$  be the auto-correlation function of a WSS process  $X(t)$ . We assume that  $R_{XX}$  is sufficiently smooth in the sense that  $R_{XX}^{(K)}(\tau)$  exists and is finite for any  $\tau \in \mathbf{R}$ , where  $K$  is a sufficiently large integer.

The next proposition expresses the auto-correlation function of the process  $X_i(t)$  and the cross-correlation function of  $X_i(t)$  and  $X(t)$  asymptotically in terms of the auto-correlation function of  $X(t)$ .

**Proposition 5.** For any  $t$  and  $\tau$  such that  $t \in [a, b]$  and  $t - \tau \in [a, b]$ ,

$$\begin{aligned} & R_{X_i X_i}(t, t - \tau) \\ &= R_{XX}(\tau) + \frac{R_{XX}^{(L)}(\tau)[(-1)^L \rho_L(2^i t) + \rho_L(2^i(t - \tau))]}{2^{iL} L!} \\ &+ \mathcal{O}(2^{-i(L+1)}) \end{aligned}$$

and

$$\begin{aligned} R_{XX_i}(t, t - \tau) &= R_{XX}(\tau) + \sum_{l=L}^{2L} \frac{R_{XX}^{(l)}(\tau) \rho_l(2^i(t - \tau))}{2^{il} l!} \\ &+ \mathcal{O}(2^{-i(L+1)}). \end{aligned}$$

Since Proposition 5 indicates that  $R_{X_i X_i}(t - \tau, t)$  depends on both  $t$  and  $\tau$ , in general the process  $X_i(t)$  is not WSS.

**Proposition 6.** The integrated MSE possesses the asymptotic form

$$e_i^2 = C_\phi \cdot 2^{-2iL} \cdot (b - a)(-1)^L R_{XX}^{(2L)}(0) + \mathcal{O}(2^{-i(2L+2)})$$

where the constant  $C_\phi$  is given by (5).

For a WSS process, the integrated MSE defined in (4) tends to infinity if it is evaluated over  $(-\infty, +\infty)$ . Therefore, it is impossible to directly extend Proposition 6 to the limiting case. However, the mean power of the approximation error is finite over  $(-\infty, +\infty)$  based on Proposition 6:

$$\begin{aligned} & \lim_{a \rightarrow -\infty, b \rightarrow +\infty} \frac{e_i^2}{b - a} \\ &= C_\phi \cdot 2^{-2iL} \cdot (-1)^L R_{XX}^{(2L)}(0) + \mathcal{O}(2^{-i(2L+2)}). \end{aligned}$$

For any stochastic process,

$$R_{XX}^{(L,L)}(t, t) = \frac{\partial^{2L} E\{X(t_1)X(t_2)\}}{\partial t_1^L \partial t_2^L} \Big|_{t_1=t_2=t}.$$

For a WSS stochastic process,

$$\begin{aligned} & \frac{\partial^{2L} E\{X(t_1)X(t_2)\}}{\partial t_1^L \partial t_2^L} \Big|_{t_1=t_2=t} \\ &= \frac{\partial^{2L} R_{XX}(t_1 - t_2)}{\partial t_1^L \partial t_2^L} \Big|_{t_1=t_2=t} \\ &= (-1)^L R_{XX}^{(2L)}(0). \end{aligned}$$

If the term  $R_{XX}^{(L,L)}(t, t)$  in Proposition 4 is replaced by  $(-1)^L R_{XX}^{(2L)}(0)$  for a WSS process  $X(t)$ , it is apparent that Proposition 4 contains Proposition 6 as a special case. However, the fact that the higher-order remainder in Proposition 6 is  $\mathcal{O}(2^{-i(2L+2)})$  instead of  $\mathcal{O}(2^{-i(2L+1)})$ , i.e., the term of the order  $2^{-i(2L+1)}$  vanishes for WSS processes, may not be directly obtained from Proposition 4.

## 6. CONCLUSIONS

We have presented coiflet-type wavelet representations of stochastic processes. Our study shows that the proposed sampling-based representations are fast, efficient, and practical. Therefore, they are promising in a large number of wavelet-based applications.

## REFERENCES

- [1] S. Cambanis and E. Masry, "Wavelet approximation of deterministic and random signals: Convergence properties and rates," *IEEE Trans. Inform. Theory*, vol. 40, pp. 1013–1029, July 1994.
- [2] D. Wei and A. C. Bovik, "Sampling approximation of smooth functions via generalized Coiflets," *IEEE Trans. Signal Processing*, vol. 46, pp. 1133–1138, Apr. 1998. Special Issue on Theory and Applications of Filter Banks and Wavelet Transforms.
- [3] I. Daubechies, "Orthonormal bases of compactly supported wavelets," *Commun. Pure Appl. Math.*, vol. 41, pp. 909–996, 1988.
- [4] I. Daubechies, *Ten Lectures on Wavelets*. Philadelphia, PA: Soc. Indus. Appl. Math., 1992.
- [5] S. G. Mallat, "A theory for multiresolution signal decomposition: the wavelet representation," *IEEE Trans. Pattern Anal. Machine Intell.*, vol. 11, pp. 674–693, July 1989.
- [6] I. Daubechies, "Orthonormal bases of compactly supported wavelets II. variations on a theme," *SIAM J. Math. Anal.*, vol. 24, pp. 499–519, Mar. 1993.
- [7] D. Wei and A. C. Bovik, "Generalized Coiflets with nonzero-centered vanishing moments," *IEEE Trans. Circuits Syst. II*, vol. 45, pp. 988–1001, Aug. 1998. Special Issue on Multirate Systems, Filter Banks, Wavelets, and Applications.

# TIME-FREQUENCY COHERENCE ANALYSIS OF NONSTATIONARY RANDOM PROCESSES\*

Gerald Matz and Franz Hlawatsch

Institute of Communications and Radio-Frequency Engineering, Vienna University of Technology  
Gusshausstrasse 25/389, A-1040 Vienna, Austria  
phone: +43 1 58801 38916, fax: +43 1 58801 38999, email: gmatz@aurora.nt.tuwien.ac.at  
web: http://www.nt.tuwien.ac.at/dspgroup/time.html

## ABSTRACT

The coherence function is extended to nonstationary random processes through introduction and investigation of a *coherence operator* and *time-frequency (TF) coherence functions*. For *underspread* nonstationary processes, it is shown that TF coherence functions are a meaningful tool for nonstationary coherence analysis and that they provide approximate TF formulations of the coherence operator.

## 1. INTRODUCTION

Consider two jointly stationary, zero-mean, real or circular complex random processes  $x(t)$  and  $y(t)$  with power spectral densities  $P_x(f)$  and  $P_y(f)$  and cross power spectral density  $P_{x,y}(f)$ . The *coherence function* [1–3]

$$\gamma_{x,y}(f) \triangleq \frac{P_{x,y}(f)}{\sqrt{P_x(f)P_y(f)}}$$

is a practically useful, normalized measure of the cross-correlation of spectral components of  $x(t)$  and  $y(t)$ . It satisfies

$$|\gamma_{x,y}(f)|^2 \leq 1, \quad (1)$$

with  $|\gamma_{x,y}(f)|^2 \equiv 0$  iff  $x(t)$  and  $y(t)$  are uncorrelated processes ( $P_{x,y}(f) = 0$ ) and

$$|\gamma_{x,y}(f)|^2 \equiv 1 \quad (2)$$

iff  $x(t)$  and  $y(t)$  are related by an invertible linear time-invariant system,  $y(t) = (k * x)(t)$ . Furthermore,  $|\gamma_{x,y}(f)|^2$  is invariant to invertible linear process transformations, i.e., for  $a(t) = (h_1 * x)(t)$  and  $b(t) = (h_2 * y)(t)$  we have

$$|\gamma_{a,b}(f)|^2 = |\gamma_{x,y}(f)|^2. \quad (3)$$

This paper extends the coherence function to nonstationary processes. Section 2 introduces and studies a *coherence operator* of nonstationary processes. Section 3 reviews some time-frequency (TF) fundamentals. Section 4 shows that for *underspread* nonstationary processes, the TF coherence function introduced in [4] is an approximate TF formulation of the coherence operator that approximately satisfies several desirable properties. Section 5 introduces a class of TF shift covariant TF coherence functions. Simulation results are presented in Section 6. We note that proofs are omitted due to lack of space; most proofs can be found in [5].

## 2. THE COHERENCE OPERATOR

Let us consider two *nonstationary*, zero-mean, real or circular complex random processes  $x(t)$  and  $y(t)$  with autocor-

relation operators  $\mathbf{R}_x$ ,  $\mathbf{R}_y$  and cross-correlation operator<sup>1</sup>  $\mathbf{R}_{x,y}$ . The coherence function is no longer defined; however, by analogy to the coherence matrix [7], we define the *coherence operator* of  $x(t)$  and  $y(t)$  as

$$\mathbf{\Gamma}_{x,y} \triangleq \mathbf{R}_x^{-1/2} \mathbf{R}_{x,y} \mathbf{R}_y^{-1/2},$$

where, e.g.,  $\mathbf{R}_x^{-1/2}$  denotes the inverse of the positive semi-definite square-root  $\mathbf{R}_x^{1/2}$  of  $\mathbf{R}_x$  [6]. Equivalently,

$$\mathbf{\Gamma}_{x,y} = \mathbf{R}_{\tilde{x},\tilde{y}},$$

where  $\tilde{x}(t) = (\mathbf{R}_x^{-1/2} x)(t)$  and  $\tilde{y}(t) = (\mathbf{R}_y^{-1/2} y)(t)$  are stationary and white with correlation  $\mathbf{R}_{\tilde{x}} = \mathbf{R}_{\tilde{y}} = \mathbf{I}$ .

If  $x(t)$  and  $y(t)$  are jointly stationary, the kernel of  $\mathbf{\Gamma}_{x,y}$  is given by  $(\mathbf{\Gamma}_{x,y})(t_1, t_2) = \tilde{\gamma}_{x,y}(t_1 - t_2)$  with  $\tilde{\gamma}_{x,y}(\tau) = \int_{-\infty}^{\infty} \gamma_{x,y}(f) e^{j2\pi f\tau} df$ . In this sense,  $\mathbf{\Gamma}_{x,y}$  is consistent with the conventional coherence function  $\gamma_{x,y}(f)$ .

**Bounds.** The coherence operator  $\mathbf{\Gamma}_{x,y}$  satisfies bounds that are analogous to (1). Specifically, the singular values [6]  $\sigma_k \geq 0$  of  $\mathbf{\Gamma}_{x,y}$  are bounded as

$$\sigma_k \leq 1.$$

The operator norm [6]  $\|\mathbf{\Gamma}_{x,y}\|_0 \triangleq \sup_{\|g\|_2=1} \|\mathbf{\Gamma}_{x,y}g\|_2$  (with  $\|g\|_2 = [\int_{-\infty}^{\infty} |g(t)|^2 dt]^{1/2}$ ) is similarly bounded as

$$\|\mathbf{\Gamma}_{x,y}\|_0 \leq 1.$$

Finally, we have the following bounds on the (non-negative) quadratic forms induced by the positive semi-definite [6] “squared” coherence operators  $\mathbf{\Gamma}_{x,y} \mathbf{\Gamma}_{x,y}^+$  or  $\mathbf{\Gamma}_{x,y}^+ \mathbf{\Gamma}_{x,y}$  (with  $\mathbf{\Gamma}_{x,y}^+$  the adjoint [6] of  $\mathbf{\Gamma}_{x,y}$ ): for any  $g(t)$  with  $\|g\|_2 = 1$ ,

$$\langle \mathbf{\Gamma}_{x,y} \mathbf{\Gamma}_{x,y}^+ g, g \rangle \leq 1, \quad \langle \mathbf{\Gamma}_{x,y}^+ \mathbf{\Gamma}_{x,y} g, g \rangle \leq 1,$$

with the inner product defined as  $\langle x, y \rangle \triangleq \int_{-\infty}^{\infty} x(t) y^*(t) dt$ . Note that  $\mathbf{\Gamma}_{x,y} = \mathbf{0}$  iff  $x(t)$  and  $y(t)$  are uncorrelated.

**Completely coherent processes.** The “squared” coherence operators equal the identity operator,

$$\mathbf{\Gamma}_{x,y} \mathbf{\Gamma}_{x,y}^+ = \mathbf{\Gamma}_{x,y}^+ \mathbf{\Gamma}_{x,y} = \mathbf{I}$$

(equivalently,  $\mathbf{\Gamma}_{x,y}$  is a *unitary* operator [6]), iff  $y(t) = (\mathbf{K}x)(t)$  with some invertible linear (generally time-varying) system  $\mathbf{K}$ . This extends property (2) to the nonstationary case.

**Linearly distorted processes.** An extension of (3) is possible only under rather restrictive assumptions. Let

<sup>1</sup>  $\mathbf{R}_{x,y}$  is the linear operator [6] with kernel  $r_{x,y}(t_1, t_2) = E\{x(t_1) y^*(t_2)\}$ ; furthermore,  $\mathbf{R}_x = \mathbf{R}_{x,x}$ .

\*This work was supported by FWF grant P11904-TEC.

$a(t) = (\mathbf{H}_1 x)(t)$  and  $b(t) = (\mathbf{H}_2 y)(t)$  with  $\mathbf{H}_1$  and  $\mathbf{H}_2$  invertible. Then  $\Gamma_{a,b} \Gamma_{a,b}^+ = \Gamma_{x,y} \Gamma_{x,y}^+$  if  $\mathbf{H}_1$  is positive definite and commutes with  $\mathbf{R}_x$ . Similarly,  $\Gamma_{a,b}^+ \Gamma_{a,b} = \Gamma_{x,y}^+ \Gamma_{x,y}$  if  $\mathbf{H}_2$  is positive definite and commutes with  $\mathbf{R}_y$ .

### 3. TIME-FREQUENCY FUNDAMENTALS

Next, we briefly review some TF representations and concepts that will be used in subsequent sections.

- The *Weyl symbol* (WS) [8–11] of a linear operator (linear time-varying system)  $\mathbf{H}$  with kernel (impulse response)  $h(t_1, t_2)$  is defined as

$$L_{\mathbf{H}}(t, f) \triangleq \int_{-\infty}^{\infty} h\left(t + \frac{\tau}{2}, t - \frac{\tau}{2}\right) e^{-j2\pi f\tau} d\tau.$$

For an *underspread* system  $\mathbf{H}$  (see below),  $L_{\mathbf{H}}(t, f)$  can be viewed as a time-varying transfer function [10, 11].

- The *spreading function* (SF) [8, 10, 11] of a linear operator (linear time-varying system)  $\mathbf{H}$ ,

$$S_{\mathbf{H}}(\tau, \nu) \triangleq \int_{-\infty}^{\infty} h\left(t + \frac{\tau}{2}, t - \frac{\tau}{2}\right) e^{-j2\pi\nu t} dt,$$

describes the distribution of time shifts by  $\tau$  and frequency shifts by  $\nu$  effected by  $\mathbf{H}$ .

- The *Wigner-Ville spectrum* (WVS) [12–15] of nonstationary random processes  $x(t)$ ,  $y(t)$  is defined as

$$\bar{W}_{x,y}(t, f) \triangleq L_{\mathbf{R}_{x,y}}(t, f), \quad \bar{W}_x(t, f) \triangleq \bar{W}_{x,x}(t, f) \in \mathbb{R}.$$

For *jointly underspread* processes  $x(t)$ ,  $y(t)$  (see below), it can be interpreted as a time-varying (cross) power spectrum.

- Another time-varying power spectrum is the *physical spectrum* [12–16]

$$\begin{aligned} \tilde{S}_{x,y}(t, f) &\triangleq \bar{W}_{x,y}(t, f) ** W_g(-t, -f) \\ \tilde{S}_x(t, f) &\triangleq \tilde{S}_{x,x}(t, f) \geq 0, \end{aligned}$$

where  $**$  denotes 2-D convolution,  $g(t)$  is an analysis window (normalized such that  $\|g\|_2 = 1$ ), and  $W_g(t, f)$  is the Wigner distribution [12, 17] of  $g(t)$ .

- The *expected ambiguity function* [14, 18]

$$\bar{A}_{x,y}(\tau, \nu) \triangleq S_{\mathbf{R}_{x,y}}(\tau, \nu), \quad \bar{A}_x(\tau, \nu) \triangleq \bar{A}_{x,x}(\tau, \nu)$$

describes the statistical correlation of process components separated in time by  $\tau$  and in frequency by  $\nu$ .

- A system/operator  $\mathbf{H}$  is *underspread* if its SF  $S_{\mathbf{H}}(\tau, \nu)$  is supported within a rectangular region  $\mathcal{G} = [-\tau_{\mathcal{G}}, \tau_{\mathcal{G}}] \times [-\nu_{\mathcal{G}}, \nu_{\mathcal{G}}]$  of area  $\sigma_{\mathcal{G}} = 4\tau_{\mathcal{G}}\nu_{\mathcal{G}} \ll 1$  [5, 10, 11]. This means that  $\mathbf{H}$  introduces only small TF shifts. Similarly, a process  $x(t)$  is underspread if its expected ambiguity function  $\bar{A}_x(\tau, \nu)$  is supported within a rectangular region  $\mathcal{G}$  of area  $\sigma_{\mathcal{G}} \ll 1$  [14, 18]. This means that  $x(t)$  features only limited TF correlations. Two processes  $x(t)$ ,  $y(t)$  are *jointly underspread* if  $\bar{A}_x(\tau, \nu)$ ,  $\bar{A}_y(\tau, \nu)$ , and  $\bar{A}_{x,y}(\tau, \nu)$  are supported within the same rectangular region of area  $\sigma_{\mathcal{G}} \ll 1$ .

### 4. A TIME-FREQUENCY COHERENCE FUNCTION

We are now ready to study a simple and intuitively appealing TF formulation of the coherence operator  $\Gamma_{x,y}$  that avoids operator inversions. A *TF coherence function* based on the WVS was defined in [4] as

$$\Gamma_{x,y}(t, f) \triangleq \frac{\bar{W}_{x,y}(t, f)}{\sqrt{\bar{W}_x(t, f) \bar{W}_y(t, f)}}, \quad (t, f) \in \mathcal{R}, \quad (4)$$

where  $\mathcal{R}$  is the TF region on which  $\bar{W}_x(t, f) > 0$  and  $\bar{W}_y(t, f) > 0$ .  $\Gamma_{x,y}(t, f)$  is a complex-valued function that is covariant to TF shifts (see Section 5) as well as to TF scalings and other metaplectic transformations of  $x(t)$ ,  $y(t)$ . For  $x(t)$ ,  $y(t)$  uncorrelated, there is  $\Gamma_{x,y}(t, f) \equiv 0$  on  $\mathcal{R}$ .

$\Gamma_{x,y}(t, f)$  as *approximate TF formulation* of  $\Gamma_{x,y}$ . We now show that for  $x(t)$ ,  $y(t)$  jointly underspread,  $\Gamma_{x,y}(t, f)$  approximates the WS of the coherence operator  $\Gamma_{x,y}$ . We start by noting that  $\Gamma_{x,y}$  can be alternatively defined by

$$\mathbf{H}_x \Gamma_{x,y} \mathbf{H}_y = \mathbf{R}_{x,y}, \quad (5)$$

with  $\mathbf{H}_x = \mathbf{R}_x^{1/2}$  and  $\mathbf{H}_y = \mathbf{R}_y^{1/2}$ . Our central assumption will be that  $S_{\mathbf{H}_x}(\tau, \nu)$ ,  $S_{\mathbf{H}_y}(\tau, \nu)$ , and  $\bar{A}_{x,y}(\tau, \nu)$  are supported within the same rectangular region  $\mathcal{G} = [-\tau_{\mathcal{G}}, \tau_{\mathcal{G}}] \times [-\nu_{\mathcal{G}}, \nu_{\mathcal{G}}]$  of area  $\sigma_{\mathcal{G}} = 4\tau_{\mathcal{G}}\nu_{\mathcal{G}}$ .

We can split the coherence operator  $\Gamma_{x,y}$  into a part  $\Gamma_{x,y}^{\mathcal{G}}$  whose SF is supported within  $\mathcal{G}$  and a part  $\Gamma_{x,y}^{\bar{\mathcal{G}}}$  whose SF is supported outside  $\mathcal{G}$ . This is motivated by the desire of approximating  $\Gamma_{x,y}$  by  $\Gamma_{x,y}^{\mathcal{G}}$  in the sense that replacing  $\Gamma_{x,y}$  by  $\Gamma_{x,y}^{\mathcal{G}}$  does not greatly affect the validity of (5):

$$\mathbf{H}_x \Gamma_{x,y} \mathbf{H}_y = \mathbf{R}_{x,y} \implies \mathbf{H}_x \Gamma_{x,y}^{\mathcal{G}} \mathbf{H}_y \approx \mathbf{R}_{x,y}. \quad (6)$$

Indeed, we have the following result.

**Theorem 1** [5]. *Under the assumption stated above, the difference  $\mathbf{H}_x \Gamma_{x,y}^{\mathcal{G}} \mathbf{H}_y - \mathbf{R}_{x,y}$  is bounded as<sup>2</sup>*

$$\frac{\|\mathbf{H}_x \Gamma_{x,y}^{\mathcal{G}} \mathbf{H}_y - \mathbf{R}_{x,y}\|_2}{\|\mathbf{H}_x\|_2 \|\Gamma_{x,y}^{\mathcal{G}}\|_2 \|\mathbf{H}_y\|_2} \leq 3\sqrt{\sigma_{\mathcal{G}}}.$$

Hence, if  $\sigma_{\mathcal{G}} \ll 1$ , i.e., if  $x(t)$  and  $y(t)$  are jointly underspread, the approximation in (6) is indeed valid.

We now pass to the TF domain using the WS.

**Theorem 2** [5]. *Under the assumption stated above, the difference  $\Delta_1(t, f) \triangleq L_{\mathbf{H}_x}(t, f) L_{\Gamma_{x,y}^{\mathcal{G}}}(t, f) L_{\mathbf{H}_y}(t, f) - \bar{W}_{x,y}(t, f)$  is bounded as<sup>3</sup>*

$$\frac{|\Delta_1(t, f)|}{\|S_{\mathbf{H}_x}\|_1 \|S_{\Gamma_{x,y}^{\mathcal{G}}}\|_{\infty} \|S_{\mathbf{H}_y}\|_1} \leq \frac{3\pi}{2} \sigma_{\mathcal{G}}^2 + 9\sigma_{\mathcal{G}}.$$

Hence, for  $\sigma_{\mathcal{G}} \ll 1$  one has

$$L_{\mathbf{H}_x}(t, f) L_{\Gamma_{x,y}^{\mathcal{G}}}(t, f) L_{\mathbf{H}_y}(t, f) \approx \bar{W}_{x,y}(t, f). \quad (7)$$

We now insert the approximations  $L_{\mathbf{H}_x}(t, f) \approx \sqrt{\bar{W}_x(t, f)}$  and  $L_{\mathbf{H}_y}(t, f) \approx \sqrt{\bar{W}_y(t, f)}$  valid for underspread  $x(t)$  and for underspread  $y(t)$  [5] and divide by  $\sqrt{\bar{W}_x(t, f) \bar{W}_y(t, f)}$  on  $\mathcal{R}$ . Equation (7) thus becomes

$$L_{\Gamma_{x,y}^{\mathcal{G}}}(t, f) \approx \Gamma_{x,y}(t, f), \quad (t, f) \in \mathcal{R}, \quad (8)$$

where  $\Gamma_{x,y}(t, f)$  is the TF coherence function in (4). Furthermore, it can be shown [5] that  $L_{\Gamma_{x,y}^{\mathcal{G}}}(t, f)$  equals  $L_{\Gamma_{x,y}}(t, f)$

<sup>2</sup>Here,  $\|\mathbf{H}\|_2 \triangleq [\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |h(t_1, t_2)|^2 dt_1 dt_2]^{1/2}$ .

<sup>3</sup>We note that  $\|S_{\mathbf{H}}\|_1 \triangleq \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |S_{\mathbf{H}}(\tau, \nu)| d\tau d\nu$  and  $\|S_{\mathbf{H}}\|_{\infty} \triangleq \sup_{\tau, \nu} |S_{\mathbf{H}}(\tau, \nu)|$ .

convolved by a function  $\psi(t, f)$  whose 2-D Fourier transform is 1 in  $\mathcal{G}$  and 0 outside  $\mathcal{G}$ . For  $\sigma_{\mathcal{G}}$  small,  $\psi(t, f)$  is a smooth function and thus  $L_{\Gamma_{x,y}}(t, f)$  is a smoothed version of  $L_{\Gamma_{x,y}}(t, f)$ . Hence, (8) states that for jointly underspread  $x(t)$ ,  $y(t)$ , the TF coherence function  $\Gamma_{x,y}(t, f)$  is approximately equal to a smoothed version of the WS of the coherence operator  $\Gamma_{x,y}$ . In this sense,  $\Gamma_{x,y}(t, f)$  provides an approximate TF formulation of the coherence operator  $\Gamma_{x,y}$ .

**Bounds.** In Section 5, it will be shown that the alternative TF coherence function

$$\Gamma'_{x,y}(t, f) \triangleq \frac{\tilde{S}_{x,y}(t, f)}{\sqrt{\tilde{S}_x(t, f) \tilde{S}_y(t, f)}} \quad (9)$$

satisfies the bound  $|\Gamma'_{x,y}(t, f)|^2 \leq 1$ . For  $x(t)$  and  $y(t)$  jointly underspread, the WVS are approximately equal to the corresponding physical spectra [5, 14, 18], and thus  $\Gamma_{x,y}(t, f) \approx \Gamma'_{x,y}(t, f)$  or, equivalently,  $\tilde{S}_x(t, f) \tilde{S}_y(t, f) |\overline{W}_{x,y}(t, f)|^2 \approx \overline{W}_x(t, f) \overline{W}_y(t, f) |\tilde{S}_{x,y}(t, f)|^2$ . The last approximation is supported by the following result.

**Theorem 3 [5].** Let  $\bar{A}_x(\tau, \nu)$ ,  $\bar{A}_y(\tau, \nu)$ , and  $\bar{A}_{x,y}(\tau, \nu)$  be supported within the same rectangle  $\mathcal{G} = [-\tau_{\mathcal{G}}, \tau_{\mathcal{G}}] \times [-\nu_{\mathcal{G}}, \nu_{\mathcal{G}}]$ . Then, the difference  $\Delta_2(t, f) \triangleq \tilde{S}_x(t, f) \tilde{S}_y(t, f) |\overline{W}_{x,y}(t, f)|^2 - \overline{W}_x(t, f) \overline{W}_y(t, f) |\tilde{S}_{x,y}(t, f)|^2$  is bounded as

$$\frac{|\Delta_2(t, f)|}{\|\bar{A}_x\|_1 \|\bar{A}_y\|_1 \|\bar{A}_{x,y}\|_1^2} \leq 4\epsilon,$$

where  $\epsilon \triangleq \max_{(\tau, \nu) \in \mathcal{G}} |1 - A_g(\tau, \nu)|$  with  $A_g(\tau, \nu)$  the ambiguity function [12, 17] of the analysis window  $g(t)$  used in the physical spectrum.

Since  $A_g(\tau, \nu) \approx 1$  for small  $(\tau, \nu)$ ,  $\epsilon$  will be small for small  $\mathcal{G}$  and thus, still for small  $\mathcal{G}$ ,

$$|\Gamma_{x,y}(t, f)|^2 \approx |\Gamma'_{x,y}(t, f)|^2. \quad (10)$$

With  $|\Gamma'_{x,y}(t, f)|^2 \leq 1$ , (10) implies that for  $x(t)$ ,  $y(t)$  jointly underspread,  $|\Gamma_{x,y}(t, f)|^2$  is approximately bounded by 1.

However, for  $x(t)$ ,  $y(t)$  not underspread,  $|\Gamma_{x,y}(t, f)|^2$  may be arbitrarily large. Consider for example the two correlated random processes  $x(t) = \beta u(t + t_0) e^{-j2\pi f_0 t}$  and  $y(t) = \beta u(t - t_0) e^{j2\pi f_0 t}$  where  $u(t) = e^{-\pi t^2/T^2} / \sqrt{2T}$ ,  $t_0$  and  $f_0$  are fixed, and  $\beta$  is random with  $E\{|\beta|^2\} = \gamma > 0$ . One obtains

$$\overline{W}_{x,y}(t, f) = 2\gamma e^{-2\pi[t^2/T^2 + f^2 T^2]} \cos(4\pi(f_0 t - t_0 f + t_0 f_0))$$

$$\overline{W}_x(t, f) = \gamma e^{-2\pi[(t+t_0)^2/T^2 + (f+f_0)^2 T^2]}$$

$$\overline{W}_y(t, f) = \gamma e^{-2\pi[(t-t_0)^2/T^2 + (f-f_0)^2 T^2]}.$$

It is seen that  $\overline{W}_x(t, f)$  and  $\overline{W}_y(t, f)$  are localized about  $(-t_0, -f_0)$  and  $(t_0, f_0)$ , respectively. However,  $\overline{W}_{x,y}(t, f)$  is localized (and oscillatory) about  $(0, 0)$ , corresponding to a "statistical cross term" [14]. It follows that

$$|\Gamma_{x,y}(0, 0)| = 2e^{2\pi[t_0^2/T^2 + f_0^2 T^2]} |\cos(4\pi t_0 f_0)|,$$

which for increasing  $t_0$ ,  $f_0$  can become arbitrarily large. This refutes a previous incorrect claim that  $|\Gamma_{x,y}(t, f)|$  is bounded by 1 [4]. Furthermore, we see that the large values of  $|\Gamma_{x,y}(t, f)|$  are due to TF correlations [5, 14, 18], i.e., correlations between components of  $x(t)$  and  $y(t)$  located in different parts of the TF plane, which give rise to "statistical

cross terms" in  $\overline{W}_{x,y}(t, f)$  [14]. We note that for large  $t_0$ ,  $f_0$ , the processes  $x(t)$  and  $y(t)$  are not jointly underspread.

**Completely coherent processes.** We next consider the case of linearly related processes  $y(t) = (\mathbf{K}x)(t)$ , where we would like to have

$$|\Gamma_{x,y}(t, f)|^2 \approx 1, \quad (t, f) \in \mathcal{R} \quad (11)$$

or, equivalently,  $|\overline{W}_{x,y}(t, f)|^2 \approx \overline{W}_x(t, f) \overline{W}_y(t, f)$ .

**Theorem 4 [5].** Let  $S_{\mathbf{K}}(\tau, \nu)$  and  $\bar{A}_x(\tau, \nu)$  be supported within the same rectangle  $\mathcal{G} = [-\tau_{\mathcal{G}}, \tau_{\mathcal{G}}] \times [-\nu_{\mathcal{G}}, \nu_{\mathcal{G}}]$  of area  $\sigma_{\mathcal{G}} = 4\tau_{\mathcal{G}}\nu_{\mathcal{G}}$ . Then, the difference  $\Delta_3(t, f) \triangleq |\overline{W}_{x,y}(t, f)|^2 - \overline{W}_x(t, f) \overline{W}_y(t, f)$  is bounded as

$$\frac{|\Delta_3(t, f)|}{\|\bar{A}_x\|_1^2 \|S_{\mathbf{K}}\|_1^2} \leq \frac{11\pi}{2} \sigma_{\mathcal{G}}.$$

Hence, for small  $\sigma_{\mathcal{G}}$ ,  $|\overline{W}_{x,y}(t, f)|^2 \approx \overline{W}_x(t, f) \overline{W}_y(t, f)$  and the approximation (11) is indeed valid. Small  $\sigma_{\mathcal{G}}$  implies that  $x(t)$  and  $\mathbf{K}$  are jointly underspread; in this case,  $y(t) = (\mathbf{K}x)(t)$  will be underspread as well. An example where  $\mathbf{K}$  is not underspread and thus (11) is not valid was given further above. Indeed, the processes  $x(t) = \beta u(t + t_0) e^{-j2\pi f_0 t}$  and  $y(t) = \beta u(t - t_0) e^{j2\pi f_0 t}$  defined above are related as  $y(t) = (\mathbf{K}x)(t)$ , where  $\mathbf{K}$  is a TF shift operator which for large  $t_0$ ,  $f_0$  is not underspread.

**Linearly distorted processes.** For  $a(t) = (\mathbf{H}_1 x)(t)$  and  $b(t) = (\mathbf{H}_2 y)(t)$ , we would like to have the (approximate) invariance

$$|\Gamma_{a,b}(t, f)|^2 \approx |\Gamma_{x,y}(t, f)|^2, \quad (t, f) \in \mathcal{R}, \quad (12)$$

which equivalently requires  $|\overline{W}_{a,b}(t, f)|^2 \overline{W}_x(t, f) \overline{W}_y(t, f) \approx |\overline{W}_{x,y}(t, f)|^2 \overline{W}_a(t, f) \overline{W}_b(t, f)$ .

**Theorem 5 [5].** Let  $S_{\mathbf{H}_1}(\tau, \nu)$ ,  $S_{\mathbf{H}_2}(\tau, \nu)$ ,  $\bar{A}_x(\tau, \nu)$ ,  $\bar{A}_y(\tau, \nu)$ , and  $\bar{A}_{x,y}(\tau, \nu)$  be supported within the same rectangle  $\mathcal{G} = [-\tau_{\mathcal{G}}, \tau_{\mathcal{G}}] \times [-\nu_{\mathcal{G}}, \nu_{\mathcal{G}}]$  of area  $\sigma_{\mathcal{G}} = 4\tau_{\mathcal{G}}\nu_{\mathcal{G}}$ . Then, the difference  $\Delta_4(t, f) \triangleq |\overline{W}_{a,b}(t, f)|^2 \overline{W}_x(t, f) \overline{W}_y(t, f) - |\overline{W}_{x,y}(t, f)|^2 \overline{W}_a(t, f) \overline{W}_b(t, f)$  is bounded as

$$\frac{|\Delta_4(t, f)|}{\|\bar{A}_x\|_1 \|\bar{A}_y\|_1 \|\bar{A}_{x,y}\|_1^2 \|S_{\mathbf{H}_1}\|_1^2 \|S_{\mathbf{H}_2}\|_1^2} \leq \frac{9\pi}{2} \sigma_{\mathcal{G}}.$$

Hence, for small  $\sigma_{\mathcal{G}}$ , (12) is indeed valid, which means that  $|\Gamma_{x,y}(t, f)|^2$  is approximately invariant to linear process transformations. Small  $\sigma_{\mathcal{G}}$  implies that the processes  $x(t)$  and  $y(t)$  and the operators  $\mathbf{H}_1$  and  $\mathbf{H}_2$  are all jointly underspread; this implies in turn that  $a(t) = (\mathbf{H}_1 x)(t)$  and  $b(t) = (\mathbf{H}_2 y)(t)$  are jointly underspread processes as well.

## 5. SHIFT-COVARIANT TIME-FREQUENCY COHERENCE FUNCTIONS

A generalization of  $\Gamma_{x,y}(t, f)$  is given by

$$\Gamma_{x,y}^{(c)}(t, f) \triangleq \frac{P_{x,y}^{(c)}(t, f)}{\sqrt{P_x^{(c)}(t, f) P_y^{(c)}(t, f)}}, \quad (t, f) \in \mathcal{R},$$

where

$$P_{x,y}^{(c)}(t, f) \triangleq \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} r_{x,y}(t_1 + t, t_2 + t) c^*(t_1, t_2) \cdot e^{-j2\pi(t_1 - t_2)f} dt_1 dt_2 \quad (13)$$



is a *TF shift covariant time-varying power spectrum* [5, 12–15] and  $\mathcal{R}$  is the TF region on which  $P_x^{(c)}(t, f) \triangleq P_{x,x}^{(c)}(t, f) > 0$  and  $P_y^{(c)}(t, f) > 0$ . We assume that the kernel function  $c(t_1, t_2)$  in (13) satisfies  $c^*(t_2, t_1) = c(t_1, t_2)$  so that  $P_x^{(c)}(t, f)$  is real-valued. Two important special cases of  $\Gamma_{x,y}^{(c)}(t, f)$  are  $\Gamma_{x,y}(t, f)$  in (4) (obtained with  $P_{x,y}^{(c)}(t, f) = \bar{W}_{x,y}(t, f)$  or  $c(t_1, t_2) = \delta(\frac{t_1+t_2}{2})$ ) and  $\Gamma'_{x,y}(t, f)$  in (9) (obtained with  $P_{x,y}^{(c)}(t, f) = \bar{S}_{x,y}(t, f)$  or  $c(t_1, t_2) = g(t_1)g^*(t_2)$ ).

The TF coherence function  $\Gamma_{x,y}^{(c)}(t, f)$  is a complex-valued function that is parameterized by  $c(t_1, t_2)$  or, equivalently, by the (self-adjoint) linear operator  $\mathbf{C}$  with kernel  $c(t_1, t_2)$ . For fixed  $c(t_1, t_2)$ ,  $\Gamma_{x,y}^{(c)}(t, f)$  is *TF shift covariant*, i.e.,

$$\Gamma_{\tilde{x},\tilde{y}}^{(c)}(t, f) = \Gamma_{x,y}^{(c)}(t - \tau, f - \nu)$$

with  $\tilde{x}(t) = x(t - \tau)e^{j2\pi\nu t}$ ,  $\tilde{y}(t) = y(t - \tau)e^{j2\pi\nu t}$ . For  $x(t), y(t)$  uncorrelated, there is  $\Gamma_{x,y}^{(c)}(t, f) \equiv 0$  on  $\mathcal{R}$ .

**Theorem 6** [5]. *There is*

$$|\Gamma_{x,y}^{(c)}(t, f)|^2 \leq 1, \quad (t, f) \in \mathcal{R}$$

for all  $x(t), y(t)$  and with nonempty  $\mathcal{R}$  iff the operator  $\mathbf{C}$  underlying  $\Gamma_{x,y}^{(c)}(t, f)$  is positive semidefinite.<sup>4</sup>

Indeed, if  $\mathbf{C}$  is positive semidefinite, then  $P_{x,y}^{(c)}(t, f)$  is a smoothed version of  $\bar{W}_{x,y}(t, f)$  [14]; this smoothing suppresses the statistical cross terms of  $\bar{W}_{x,y}(t, f)$  present in the overspread case and thus allows  $\Gamma_{x,y}^{(c)}(t, f)$  to be properly bounded even for overspread processes.

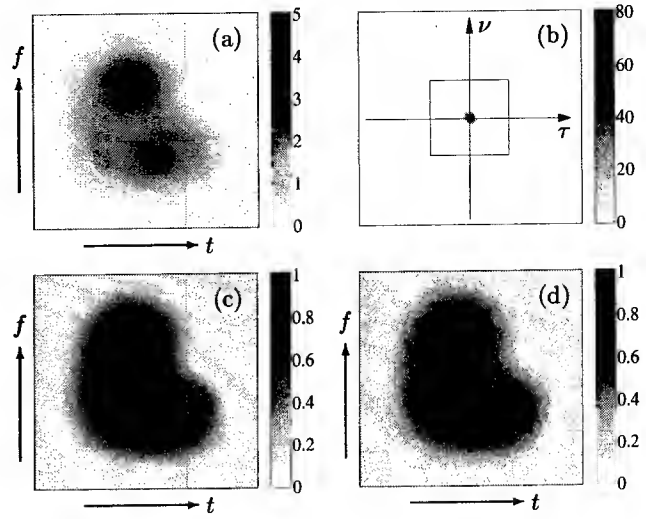
The operator  $\mathbf{C}$  underlying  $\Gamma'_{x,y}(t, f)$  in (9) is positive semidefinite, and thus  $|\Gamma'_{x,y}(t, f)|^2 \leq 1$ . On the other hand, the operator  $\mathbf{C}$  underlying  $\Gamma_{x,y}(t, f)$  is *not* positive semidefinite, and indeed we have observed in Section 4 that  $|\Gamma_{x,y}(t, f)|^2$  can be arbitrarily large (however, we recall that it is approximately bounded by 1 in the underspread case).

## 6. SIMULATION RESULTS

**Experiment 1.** We analyze the coherence of the input  $x(t)$  and the noise-contaminated output  $y(t) = (\mathbf{K}x)(t) + n(t)$  of a time-varying linear system  $\mathbf{K}$ . The input  $x(t)$  is stationary and white with correlation  $\mathbf{R}_x = \mathbf{I}$  (corresponding to constant WVS  $\bar{W}_x(t, f) \equiv 1$ ). The noise  $n(t)$  is stationary and white with correlation  $\mathbf{R}_n = \eta \mathbf{I}$  (corresponding to constant WVS  $\bar{W}_n(t, f) \equiv \eta$ ) and uncorrelated with  $x(t)$ . The WS and SF of  $\mathbf{K}$  are depicted in Figs. 1(a) and (b), respectively. The SF of  $\mathbf{K}$  shows that  $\mathbf{K}$  is underspread.

In the noise-free case ( $\eta = 0$ ),  $x(t)$  and  $y(t)$  are completely coherent, i.e.,  $\Gamma_{x,y}\Gamma_{x,y}^+ = \Gamma_{x,y}^+\Gamma_{x,y} = \mathbf{I}$ . Since  $x(t)$  and  $\mathbf{K}$  are underspread, Theorem 4 applies and we can expect that  $|\Gamma_{x,y}(t, f)|^2 \approx 1$ . Indeed, we found that the maximum deviation of  $|\Gamma_{x,y}(t, f)|^2$  from 1 was 0.028.

For  $\eta > 0$ , the noise causes a reduction of coherence that depends on the output SNR. Since the output SNR is TF-dependent (due to the TF weighting characteristic of  $\mathbf{K}$  as shown in Fig. 1(a)), the coherence reduction is TF-dependent as well. This is clearly indicated by the WS of



**Figure 1:** Simulation results for Experiment 1: (a) WS of  $\mathbf{K}$ ; (b) magnitude of SF of  $\mathbf{K}$ ; (c) magnitude of WS of  $\Gamma'_{x,y}$ ; (d) magnitude of  $\Gamma_{x,y}(t, f)$ . The rectangle in (b) has area 1 and allows to assess the underspread property of  $\mathbf{K}$ . Time duration is 256 samples; normalized frequency ranges from  $-1/4$  to  $1/4$ .

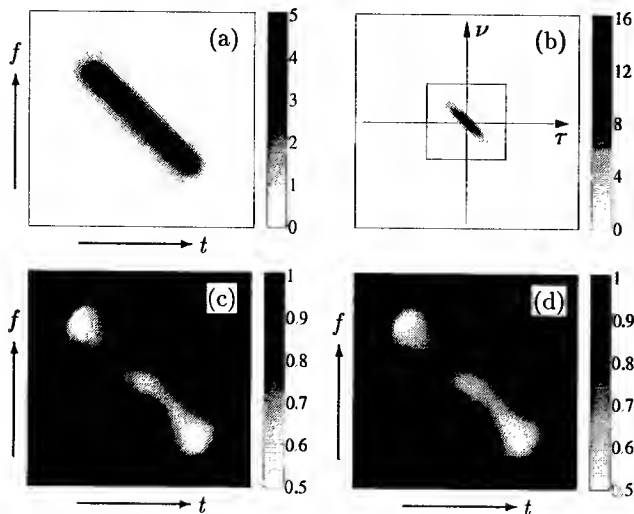
$\Gamma_{x,y}^G$  and the TF coherence function  $\Gamma_{x,y}(t, f)$  shown in Figs. 1(c) and (d), respectively. Moreover, the similarity of these two results confirms the validity of the approximation (8).

**Experiment 2.** Again,  $y(t) = (\mathbf{K}x)(t) + n(t)$  with  $x(t)$  and  $\mathbf{K}$  as in the previous example. However,  $n(t)$  now is nonstationary narrowband noise with WVS as shown in Fig. 2(a). From the expected ambiguity function of  $n(t)$  shown in Fig. 2(b), it is seen that  $n(t)$  is reasonably underspread. The Weyl symbol of  $\Gamma_{x,y}^G$  and the TF coherence function  $\Gamma_{x,y}(t, f)$ , shown respectively in Figs. 2(c) and (d), are again seen to be practically identical. In this example, significant coherence reduction occurs only in the TF support region of the noise; in the remainder of the TF plane there is complete coherence, thus indicating a pure linear relation between those components of  $x(t)$  and  $y(t)$  that are located in this “noise-free” TF region. Again, both  $L_{\Gamma_{x,y}^G}(t, f)$  and  $\Gamma_{x,y}(t, f)$  clearly indicate the TF dependence of coherence.

**Experiment 3.** We finally analyze the coherence of pressure signals  $x(t)$  measured inside the cylinder of a combustion engine and vibration signals  $y(t)$  measured on the engine block.<sup>5</sup> The goal is to see whether the pressure and vibration processes are linearly related (as assumed in [19]). Both  $x(t)$  and  $y(t)$  consist of several resonances with decreasing resonance frequencies. Estimates  $\hat{\Gamma}_{x,y}(t, f)$  of the TF coherence function  $\Gamma_{x,y}(t, f)$  are shown in Fig. 3 for two different engine speeds. (These estimates were computed using estimated Wigner-Ville spectra [20] obtained from multiple realizations.) For both engine speeds,  $|\hat{\Gamma}_{x,y}(t, f)|$  is seen to be significantly larger than zero in the TF support regions of the resonances. Specifically, in the TF region of

<sup>4</sup>We recall that a positive semidefinite operator  $\mathbf{C}$  is defined by the condition  $\langle \mathbf{C}x, x \rangle \geq 0$  for all  $x(t)$  [6]. For  $\mathbf{C}$  positive semidefinite, there is  $P_x^{(c)}(t, f) \geq 0$  for all  $(t, f)$  and for all  $x(t)$ .

<sup>5</sup>We are grateful to S. Carstens-Behrens, M. Wagner, and J. F. Böhme for providing us with the car engine data (courtesy of Aral-Forschung, Bochum).



**Figure 2:** Simulation results for Experiment 2: (a) WVS of  $n(t)$ ; (b) magnitude of expected ambiguity function of  $n(t)$ ; (c) magnitude of WS of  $\Gamma_{x,y}^G$ ; (d) magnitude of  $\Gamma_{x,y}(t, f)$ . The rectangle in (b) has area 1 and allows to assess the underspread property of  $n(t)$ . Time duration is 256 samples; normalized frequency ranges from  $-1/4$  to  $1/4$ .

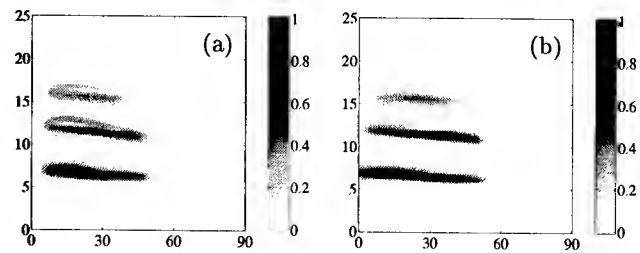
the first resonance the maximum of  $|\hat{\Gamma}_{x,y}(t, f)|$  is about 0.9, which clearly indicates a linear relationship. For the second and third resonance, the maximum of  $|\hat{\Gamma}_{x,y}(t, f)|$  is about 0.7 and 0.4, respectively. This still suggests a linear relationship, though apparently contaminated by measurement noise and extraneous interference.

## 7. CONCLUSIONS

We introduced and studied a coherence operator and time-frequency (TF) coherence functions for nonstationary coherence analysis. We showed that for jointly underspread nonstationary processes, TF coherence functions are meaningful tools for nonstationary coherence analysis. However, if the processes are not jointly underspread, meaningful results can only be obtained with TF coherence functions based on smoothed time-varying spectra. We note that TF coherence functions can be estimated based on estimates of the time-varying spectra involved [4, 12, 13, 20]. Furthermore, many of the theorems presented can be extended to a generalized underspread concept that does not require exact compact support of spreading functions and expected ambiguity functions [5, 10, 14].

## REFERENCES

- [1] J. S. Bendat and A. G. Piersol, *Engineering Applications of Correlation and Spectral Analysis*. New York: Wiley, 2nd ed., 1993.
- [2] J. A. Cadzow and O. M. Solomon, "Linear modeling and the coherence function," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 35, pp. 19–28, Jan. 1987.
- [3] W. A. Gardner, *Statistical Spectral Analysis*. New Jersey: Prentice Hall, 1988.
- [4] L. B. White and B. Boashash, "Cross spectral analysis of nonstationary processes," *IEEE Trans. Inf. Theory*, vol. 36, pp. 830–835, July 1990.



**Figure 3:** Simulation results for Experiment 3: Magnitude of estimated TF coherence function  $\hat{\Gamma}_{x,y}(t, f)$  at (a) 2000 rpm and (b) 3000 rpm. Horizontal axis: crank angle in degrees (proportional to time); vertical axis: frequency in kHz.

- [5] G. Matz, *A Time-Frequency Calculus for Underspread Systems and Processes with Applications*. PhD thesis, Vienna University of Technology, in preparation.
- [6] A. W. Naylor and G. R. Sell, *Linear Operator Theory in Engineering and Science*. New York: Springer, 2nd ed., 1982.
- [7] L. L. Scharf and J. B. Thomas, "Wiener filters in canonical coordinates for transform coding, filtering, and quantizing," *IEEE Trans. Signal Processing*, vol. 46, pp. 647–654, March 1998.
- [8] W. Kozek, "Time-frequency signal processing based on the Wigner-Weyl framework," *Signal Processing*, vol. 29, pp. 77–92, Oct. 1992.
- [9] R. G. Shenoy and T. W. Parks, "The Weyl correspondence and time-frequency analysis," *IEEE Trans. Signal Processing*, vol. 42, pp. 318–331, Feb. 1994.
- [10] G. Matz and F. Hlawatsch, "Time-frequency transfer function calculus (symbolic calculus) of linear time-varying systems (linear operators) based on a generalized underspread theory," *J. Math. Phys.*, vol. 39, pp. 4041–4071, Aug. 1998.
- [11] W. Kozek, "On the transfer function calculus for underspread LTV channels," *IEEE Trans. Signal Processing*, vol. 45, pp. 219–223, Jan. 1997.
- [12] P. Flandrin, *Time-Frequency/Time-Scale Analysis*. San Diego (CA): Academic Press, 1999.
- [13] P. Flandrin and W. Martin, "The Wigner-Ville spectrum of nonstationary random signals," in *The Wigner Distribution — Theory and Applications in Signal Processing* (W. Mecklenbräuker and F. Hlawatsch, eds.), pp. 211–267, Amsterdam (The Netherlands): Elsevier, 1997.
- [14] G. Matz and F. Hlawatsch, "Time-varying spectra for underspread and overspread nonstationary processes," in *Proc. 32nd Asilomar Conf. Signals, Systems, Computers*, (Pacific Grove, CA), pp. 282–286, Nov. 1998.
- [15] M. Amin, "Time-frequency spectrum analysis and estimation for non-stationary random processes," in *Advances in Spectrum Estimation* (B. Boashash, ed.), pp. 208–232, Melbourne: Longman Cheshire, 1992.
- [16] W. D. Mark, "Spectral analysis of the convolution and filtering of non-stationary stochastic processes," *J. Sound Vib.*, vol. 11, no. 1, pp. 19–63, 1970.
- [17] W. Mecklenbräuker and F. Hlawatsch, eds., *The Wigner Distribution — Theory and Applications in Signal Processing*. Amsterdam (The Netherlands): Elsevier, 1997.
- [18] W. Kozek, F. Hlawatsch, H. Kirchauer, and U. Trautwein, "Correlative time-frequency analysis and classification of nonstationary random processes," in *Proc. IEEE-SP Int. Sympos. Time-Frequency Time-Scale Analysis*, (Philadelphia, PA), pp. 417–420, Oct. 1994.
- [19] S. Carstens-Behrens, M. Wagner, and J. F. Böhme, "Improved knock detection by time-variant filtered structure-borne sound," in *Proc. IEEE ICASSP-99*, (Phoenix, AZ), pp. 2255–2258, March 1999.
- [20] W. Kozek and K. Riedel, "Quadratic time-varying spectral estimation for underspread processes," in *Proc. IEEE-SP Int. Sympos. Time-Frequency Time-Scale Analysis*, (Philadelphia, PA), pp. 460–463, Oct. 1994.

# MULTI-COMPONENT IF ESTIMATION

*Zahir M. Hussain and Boualem Boashash*

Signal Processing Research Centre, Queensland University of Technology  
Brisbane, Q.4000, Australia  
e-mail: z.hussain@qut.edu.au, b.boashash@qut.edu.au

## ABSTRACT

An adaptive approach to the estimation of the instantaneous frequency (IF) of non-stationary mono- and multi-component FM signals with additive Gaussian noise is presented. It is shown that the bias and variance of the IF estimate are functions of the lag window length. If there is a bias-variance tradeoff, then the optimal window length for this tradeoff depends on the unknown IF law. Hence an adaptive algorithm with a time-varying and data-driven window length is needed. The adaptive algorithm can utilize any quadratic time-frequency distribution that satisfies certain conditions. A quadratic distribution that is most suitable for this approach is proposed. The algorithm estimates multiple IF laws by using a tracking algorithm for the signal components and utilizing the property that the proposed distribution enables non-parametric component amplitudes estimation. An extension of the proposed TFD consisting in the use of time-only kernels for adaptive IF estimation is also proposed.

## 1. INTRODUCTION

The problem of non-parametric instantaneous frequency (IF) estimation for multi-component non-stationary signals is an important and unresolved issue in signal processing. Time-frequency analysis techniques are generally used as they reveal the multi-component nature of such signals.

The concept of instantaneous frequency and methods of IF estimation were reviewed in [1] and [2]. An efficient adaptive algorithm for IF estimation using the Wigner-Ville distribution (WVD) was presented in [3] and [4]. This paper aims to develop a general adaptive method for IF estimation of mono- and multi-component signals in additive Gaussian noise that is suitable for quadratic time-frequency distributions. We found that, to be used for this purpose, the quadratic TFDs must satisfy three conditions:

- first, the variance of the IF estimate using a TFD  $\rho(t, f)$  should be a continuously decreasing function of the lag window length while the bias is continuously increasing, so that the algorithm will converge at the optimal lag window length that resolves this bias-variance tradeoff.
- second, since we introduce an adaptive window length in the lag direction, the kernel of  $\rho(t, f)$  should not have a narrow passband in the lag direction which would limit the effective length of the adaptive lag window.
- third,  $\rho(t, f)$  should have a high time-frequency resolution while suppressing cross-terms efficiently so as to give

a robust IF estimate for mono- and multi-component FM signals.

In this analysis we propose a distribution  $d(t, f)$  that is most suitable for the adaptive IF algorithm in the sense that it has high resolution, effective cross-terms reduction, and a kernel that does not perform filtering in the lag direction; in addition, it enables non-parametric amplitude estimation.

## 2. A HIGH-RESOLUTION TFD

### 2.1. The Time-Lag Kernel

Recently a time-frequency distribution  $B(t, f)$  was proposed and shown to be superior to other fixed-kernel TFDs in terms of cross-terms reduction and resolution enhancement [5]. We have used this distribution for IF estimation for mono- and multi-component FM signals [6]. However, no direct component amplitudes estimation is possible from  $B(t, f)$  or other quadratic TFDs, a difficulty that appears in the case of adaptive IF estimation of multi-component signals. Based on  $B(t, f)$  and the conditions of the adaptive algorithm, the kernel of the proposed distribution  $d(t, f)$  in the time-lag domain is given by

$$G(t, \tau) = G_{\alpha}(t) = \frac{k_{\alpha}}{\cosh^{2\alpha}(t)} \quad (1)$$

where  $\alpha$  is a real positive number and  $k_{\alpha} = \Gamma(2\alpha)/\Gamma^2(\alpha)$ ,  $\Gamma$  stands for the gamma function. Filtering in the  $\tau$  direction is performed by introducing a window function (see section III).

### 2.2. Properties of the Proposed Distribution

Most of the desirable properties of time-frequency distributions explained in [1] and [2] are satisfied by this kernel as stated below.

- Realness, time-shift and frequency shift invariance, frequency marginal and group delay, and the frequency support properties are satisfied. The time support property is not strictly satisfied, but it is approximately true.
- Reduced interference and resolution: This property is satisfied by  $d(t, f)$ . First we consider the sum of two complex sinusoidal signals  $z(t) = z_1(t) + z_2(t) = a_1 e^{j(2\pi f_1 t + \theta_1)} + a_2 e^{j(2\pi f_2 t + \theta_2)}$  where  $a_1, a_2, \theta_1$  and  $\theta_2$  are constants. The time-frequency distribution  $d(t, f)$  of the signal  $z(t)$  is ob-

tained as [1]

$$d(t, f) = a_1^2 \delta(f - f_1) + a_2^2 \delta(f - f_2) + 2a_1 a_2 \gamma_\alpha(t) \delta\left(f - \frac{f_1 + f_2}{2}\right) \quad (2)$$

where

$$\gamma_\alpha(t) = \frac{|\Gamma(\alpha + j\pi(f_1 - f_2))|^2}{\Gamma^2(\alpha)} \cos(2\pi(f_1 - f_2)t + \theta_1 - \theta_2).$$

It is clear that the cross-terms are oscillatory in time and depend on the frequency separation between signal components. If  $f_1$  and  $f_2$  are well separated then the term  $|\Gamma(\alpha + j\pi(f_1 - f_2))|^2$  can be substantially reduced, while  $\Gamma^2(\alpha)$  can be made high if  $\alpha$  is small. As  $\alpha \rightarrow \infty$ , we have  $G(t, \tau) \rightarrow \delta(t)$  and  $d(t, f)$  would approach the Wigner-Ville distribution  $W(t, f)$ .

For FM signals,  $d(t, f)$  performs well in reducing interference (cross-terms) while keeping high resolution. Figure 1 shows a comparison between the discrete versions of  $d(t, f)$  and the Choi-Williams distribution  $CW(t, f)$  using a two-component linear FM signal. Changing the parameters  $\sigma$  and  $\alpha$  will improve one of the above two requirements at the expense of the other.

• Time marginal and instantaneous frequency: Now we consider the important property of the instantaneous frequency of a time-varying signal  $s(t)$ . The instantaneous frequency  $f_i(t)$  is defined as  $f_i(t) = \frac{1}{2\pi} \frac{d\phi(t)}{dt}$  where  $z(t) = a(t)e^{j\phi(t)}$  is the analytic signal associated with  $s(t)$  [1]. Traditionally a time-frequency distribution  $\rho(t, f)$  is looked upon as analogous to a probability distribution, hence it was imposed that the first moment of  $\rho(t, f)$  with respect to  $f$  must equal the instantaneous frequency  $f_i(t)$ , leading to the conditions  $g(\nu, 0) = 1 \forall \nu$  and  $\partial g(\nu, \tau)/\partial \tau|_{\tau=0} = 0 \forall \nu$ , where  $g(\nu, \tau) = \int_{-\infty}^{\infty} \rho(t, f) e^{-j2\pi f \tau} df$  [1]. But the spectrogram, Page, and Rihaczek distributions, for example, do not satisfy these conditions [1]. If the time-frequency distribution does not satisfy one of the marginals, the analogy with a probability distribution breaks down and the traditional reasoning for the IF property is no longer valid. Hence we postulate the following general IF property: at any time  $t$ , the time-frequency distribution  $\rho(t, f)$  should have absolute maximum at  $f = \frac{1}{2\pi} \frac{d\phi(t)}{dt}$ , which is the actual important characteristic needed for IF estimation.

In fact,  $d(t, f)$  does not satisfy the time marginal, hence does not satisfy the traditional condition for the instantaneous frequency. But we shall prove that at any  $t$ ,  $d(t, f)$  has an absolute maximum at  $f = \frac{1}{2\pi} \frac{d\phi(t)}{dt}$  for linear FM. This is the basis for our IF estimate. For non-linear FM signals this estimate is biased. This bias would be the basis of the adaptive IF estimation developed in section III.

□ Proof: For an FM signal of the form  $z(t) = a e^{j\phi(t)}$ , we can express  $d(t, f)$  as [1]

$$\begin{aligned} d(t, f) &= |a|^2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-j2\pi f \tau} G(t - u, \tau) \\ &\quad \times e^{j[\phi(u + \tau/2) - \phi(u - \tau/2)]} du d\tau \\ &= |a|^2 \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} e^{-j2\pi f \tau} G(t - u, \tau) \\ &\quad \times e^{j\tau[\phi'(u) + \sum_{k=3}^{\infty} \frac{\tau^{k-1}}{k!2^{k-1}} \phi^{(k)}(u)]} du d\tau \end{aligned}$$

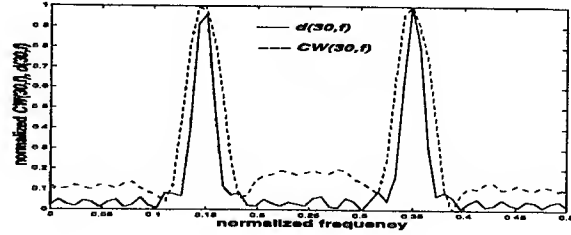


Figure 1: Performance comparison between  $d(t, f)$  for  $\alpha = 0.1$  and  $CW(t, f)$  for  $\sigma = 11$  using a two-component linear FM signal at the discrete time instant  $n = 30$ . Total signal length is  $N = 64$  and the sampling interval is  $T = 1$ .

using Taylor series expansion. Assuming a relatively small effect from higher-order derivatives  $\phi^{(k)}(t)$ ,  $k \geq 3$ , we have

$$\begin{aligned} d(t, f) &\approx |a|^2 \int_{-\infty}^{\infty} G_\alpha(t - u) \delta\left[\frac{1}{2\pi} \phi'(u) - f\right] du \\ &= |a|^2 G_\alpha(t - \psi(f)) \psi'(f) \end{aligned} \quad (3)$$

where  $\psi$  is the inverse of  $\frac{1}{2\pi} \phi'$ , i.e.,  $\frac{1}{2\pi} \phi'(\psi(f)) = f$ . Assuming that  $\psi'(f)$  is not a highly peaked function of  $f$  and knowing that  $G_\alpha(t - \psi(f))$  is peaked at  $t = \psi(f)$ , the absolute maximum of  $d(t, f)$  for any time  $t$  would be at  $\psi(f) = t$ , or  $f = \frac{1}{2\pi} \phi'(t)$ , which is the instantaneous frequency of the FM signal  $z(t)$ . For non-linear FM signals, the energy peak of  $d(t, f)$  is actually biased from the instantaneous frequency because of the extra term  $\sum_{k=3}^{\infty} \frac{\tau^{k-1}}{k!2^{k-1}} \phi^{(k)}(u)$ . The major contribution in this term is due to  $\phi^{(3)}(u)$  (see section III). Therefore at the instants of rapid change in the IF law the bias is not negligible and eq.(3) would not be an accurate approximation to  $d(t, f)$  unless suitable windowing in the lag direction is used.

For linear FM signals we have  $\phi^{(k)}(t) = 0$  for  $k \geq 3$ . Assuming  $z(t) = a e^{j2\pi(f_0 t + \frac{\beta_0}{2} t^2)}$ , where  $f_0$  and  $\beta_0$  are constants, we have

$$d(t, f) = \frac{1}{\beta_0} |a|^2 G_\alpha\left(t - \frac{1}{\beta_0}(f - f_0)\right) \quad (4)$$

which has an absolute maximum at  $f = f_0 + \beta_0 t$ , the instantaneous frequency of the linear FM signal  $z(t)$ . As  $\beta_0 \rightarrow 0$ , i.e.,  $z(t)$  approaches a sinusoid, we have  $d(t, f) \rightarrow |a|^2 \delta(f - f_0)$ , in accordance with eq.(2).

In practical implementation a window  $w(\tau)$  is used in the  $\tau$  direction and the results in eqs.(3) and (4) are convolved with the Fourier transform of  $w(\tau)$ . □

In the next section we will present an adaptive approach to the IF estimation for FM signals using quadratic time-frequency distributions.

### 3. IF ESTIMATION USING QUADRATIC TFDs

#### 3.1. Introduction to IF Estimation

We consider an analytic signal  $z(t)$  of the form

$$z(t) = a e^{j\phi(t)} + \epsilon(t)$$

where the amplitude  $a$  is constant, and  $\epsilon(t)$  is a complex-valued white Gaussian noise with independent identically distributed (i.i.d.) real and imaginary parts with total variance  $\sigma_\epsilon^2$ . The instantaneous frequency of  $z(t)$  is given by [1]

$$f_i(t) = \frac{1}{2\pi} \frac{d\phi(t)}{dt} \quad (5)$$

We assume in this analysis that  $f_i(t)$  is an arbitrary, smooth and differentiable function of time with bounded derivatives of all orders. The general equation for quadratic time-frequency representation of a signal  $z(t)$  is given by [1]

$$\rho(t, f) = \mathcal{F}_{\tau \rightarrow f} [G(t, \tau) *_{(t)} K_z(t, \tau)]$$

where  $G(t, \tau)$  is the time-lag kernel,  $K_z(t, \tau) = z(t + \frac{\tau}{2})z^*(t - \frac{\tau}{2})$  and  $*$  denotes time convolution. For smoothing and localization we introduce a window function  $w_h(\tau) = \frac{T}{h} w(\frac{\tau}{h})$  where  $w(t)$  is a real-valued symmetric window with unity length, i.e.,  $w(t) = 0$  for  $|t| > \frac{1}{2}$ ; hence the window length is  $h$ .

As the time-frequency representation is now dependent on  $h$ , we denote it by  $\rho_h(t, f)$  which is given by

$$\rho_h(t, f) = \mathcal{F}_{\tau \rightarrow f} [G_{\text{eff}}(t, \tau) *_{(t)} K_z(t, \tau)] \quad (6)$$

where  $G_{\text{eff}}$  is the effective time-lag kernel given by

$$G_{\text{eff}}(t, \tau) = w_h(\frac{\tau}{2}) G(t, \tau) \quad (7)$$

The lag window  $w_h(\frac{\tau}{2})$  will restrict the lag function of the kernel  $G(t, \tau)$  to the interval  $|\tau| < h$ . If the lag function of the kernel has a passband narrower than that of the lag window, it will dominate over the function of this window. In section III we shall prove that for a robust IF estimation,  $w_h(\tau)$  should be adaptive with variable length  $h$ . If the kernel  $G(t, \tau)$  already has a factor controlling the lag passband independently of time, it may be better to consider adapting this factor instead of introducing an adaptive window in the lag direction. However, this would require different analysis for different TFDs. In addition, there is the problem of component amplitude estimation in the case of multi-component signals. On the other hand, designing new time-lag kernels that are functions of time only could result in very efficient TFDs like  $d(t, f)$ . Such TFDs are more suitable for adaptive IF estimation as they enable non-parametric amplitude estimation. Further studies on these TFDs would be attempted in future works. In this paper we assume that the parameters of the TFD are arranged such that the lag passband of the kernel  $G(t, \tau)$  is larger than the largest lag window length necessary for the adaptive IF estimation.

In the discrete lag domain  $\rho_h(t, f)$  can be expressed as follows

$$\rho_h(t, f) = \sum_{m=-\infty}^{\infty} K_z(u, 2mT) G_{\text{eff}}(t - u, 2mT) \times e^{-j4\pi f m T} du \quad (8)$$

where  $m$  is an integer and  $T$  is the sampling interval. If  $\rho_h(t, f)$  is discretized over time and frequency then we have

$$\rho_h(n, k) = \sum_{l=-N_s}^{N_s-1} \sum_{m=-N_s}^{N_s-1} K_z(lT, 2mT) G_{\text{eff}}(nT - lT, 2mT) \times e^{-j2\pi \frac{km}{2N_s}} \quad (9)$$

where  $2N_s$  is the number of samples.

The IF estimate is a solution of the following optimization

$$\hat{f}_{ih}(t) = \arg[\max_f \rho_h(t, f)] ; 0 \leq f \leq f_s/2 \quad (10)$$

where  $f_s = 1/T$  is the sampling frequency.

### 3.2. Bias and Variance of the IF Estimate

Following the same analysis as in [6], the estimation bias is found to be

$$E[\Delta \hat{f}_{ih}(t)] = \frac{L_h(t)}{2F_h} \quad (11)$$

and the variance is

$$\text{var}(\Delta \hat{f}_{ih}(t)) = \frac{\sigma_\epsilon^2}{2|a|^2} [1 + \frac{\sigma_\epsilon^2}{2|a|^2}] \frac{E_h}{F_h^2} \quad (12)$$

where

$$\begin{aligned} \Delta \hat{f}_{ih}(t) &= \frac{1}{2\pi} \phi'(t) - \hat{f}_{ih}(t) \\ F_h &= \int_{-\infty}^{\infty} \sum_{m=-\infty}^{\infty} w_h(mT) (2\pi mT)^2 G(u, 2mT) du \\ L_h(t) &= \int_{-\infty}^{\infty} \sum_{m=-\infty}^{\infty} w_h(mT) \Delta \phi(u, mT) (2\pi mT) \\ &\quad \times G(t - u, 2mT) du \\ E_h &= \int_{-\infty}^{\infty} \sum_{m=-\infty}^{\infty} w_h(mT)^2 (2\pi mT)^2 G(u, 2mT) du \end{aligned} \quad (13)$$

Equations (11)-(13) indicate that the bias and the variance of the estimate depend on the lag window length  $h$  for any kernel  $G(t, \tau)$ . To see how the bias and the variance vary with  $h$ , asymptotic analysis as  $T \rightarrow 0$  is necessary.

### 3.3. Asymptotic Formulas Using $d(t, f)$

Following the same analysis as in [6], we have the following asymptotic formulae for the variance and the bias as  $T \rightarrow 0$  using a rectangular lag window

$$\text{var}(\Delta \hat{f}_{ih}(t)) = \frac{3\sigma_\epsilon^2}{2\pi^2 |a|^2} [1 + \frac{\sigma_\epsilon^2}{2|a|^2}] \frac{T}{h^3} \quad (14)$$

and

$$E(\Delta \hat{f}_{ih}(t)) = \frac{h^2}{80} \int_{-\infty}^{\infty} \frac{\lambda(u) du}{\cosh^{2\alpha}(t - u)} \quad (15)$$

$$E(\Delta \hat{f}_{ih}(t)) \leq \frac{M_2}{40} h^2 \quad (16)$$

where  $\lambda(t) = f_i^{(2)}(t + \tau_1) + f_i^{(2)}(t - \tau_1)$ ,  $\sup |f_i^{(2)}(t)| \leq M_2$ . For small  $h$ , the optimal window length that minimizes the mean squared error is obtained as

$$h_{\text{opt}}(t) = \left[ \frac{1800\sigma_e^2 T(1 + \frac{\sigma_e^2}{2|a|^2})}{\pi^2 |a|^2 (f_i^{(2)}(t) * 1/\cosh^{2\alpha}(t))^2} \right]^{\frac{1}{2}} \quad (17)$$

Hence the optimal window length depends on the second derivative of the instantaneous frequency  $f_i^{(2)}(t)$ , which is time and signal dependent. From eqs.(21), (25) and (26) it is clear that the variance and bias of the IF estimate using  $d(t, f)$  have the same rates of change with respect to the window length  $h$  as those using WVD [3, 4].

### 3.4. The Adaptive Algorithm and Its Conditions of Applicability

For  $d(t, f)$ , eqs.(14) and (15) show that when  $h$  increases the bias increases and the variance decreases. From eq.(17) it can be seen that the optimal window length is a function of time and depends on the second derivative of the IF law  $f_i^{(2)}(t)$ ; it decreases when the IF law  $f_i(t)$  has a high variation. Hence a time-varying window length is needed to optimize the estimation. The Stankovic-Katkovnik adaptive algorithm developed in [3] and [4] can be used for  $d(t, f)$ . In fact, this adaptive algorithm is applicable to any quadratic time-frequency distribution whose IF estimation variance is a continuously decreasing function of  $h$  while its bias is continuously increasing. These conditions are necessary for bias-variance tradeoff such that the algorithm converges at the optimum window length that resolves this tradeoff. Also the time-lag kernel of the distribution should not perform narrowband filtering in the lag direction so as not to interfere with the adaptive windowing in that direction.

The following estimates for the amplitude of the signal  $a$  and the variance of noise  $\sigma_e^2$  are used in this algorithm [4]

$$\hat{a}^2 + \hat{\sigma}_e^2 = \frac{1}{N} \sum_{n=1}^N |z(nT)|^2 \quad (18)$$

$$\hat{\sigma}_e^2 = \frac{1}{2N} \sum_{n=2}^N |z(nT) - z((n-1)T)|^2 \quad (19)$$

where  $N = 2N_s$  is the number of samples. We consider an increasing sequence of window lengths  $\{h_r \mid r = 1 : J\}$ . Since the optimal window length is time-dependent, the optimal IF estimate (as given by eq.(11)) is also time dependent. For details see [4, 6]

It should be emphasized that the above amplitude estimation works only for mono-component FM signals.

### 4. IF ESTIMATION OF MULTI-COMPONENT SIGNALS

In this section we consider a multi-component analytic signal of the form

$$z(t) = \sum_{q=1}^M (a_q e^{j\phi_q(t)} + \epsilon_q(t)) = \sum_{q=1}^M a_q e^{j\phi_q(t)} + \epsilon(t) \quad (20)$$

where the amplitudes  $\{a_q\}$  are constant,  $\epsilon_q(t)$  and  $\epsilon(t)$  are complex-valued white Gaussian noise processes with i.i.d. real and imaginary parts with total variance  $\sigma^2$  and  $\sigma_e^2 = M\sigma^2$ , respectively. The signal-to-noise ratio  $SNR$  is defined using the overall average amplitude and the overall noise. The individual IF laws for each component are given by [1]

$$f_{i,q}(t) = \frac{1}{2\pi} \frac{d\phi_q}{dt} \quad ; \quad q = 1, \dots, M. \quad (21)$$

The adaptive algorithm that tracks component maxima in the time-frequency plane requires a threshold  $\rho_{TH}(t)$  so as to ignore the local maxima caused by the cross-terms and windowing. In fact,  $\rho_{TH}(t)$  is application and distribution dependent.

The algorithm also requires the knowledge of the confidence intervals  $D_{r,q}$  for each component. The calculation of  $D_{r,q}$  depends on the estimation of the individual amplitudes  $a_i$  of the components. First we have [6]

$$\sum_{q=1}^M |\hat{a}_q|^2 + \hat{\sigma}_e^2 = \frac{1}{N} \sum_{n=1}^N |z(nT)|^2 \quad (22)$$

where  $N$  is the number of samples and the estimate of  $\hat{\sigma}_e^2$  is given by eq.(19). Hence  $\hat{\sigma}^2 = \hat{\sigma}_e^2/M$  and  $s_o = \sum_{q=1}^M |\hat{a}_q|^2$  can be calculated.

Now if the ratio between the component amplitudes can be estimated, the actual amplitudes can be estimated. If we assume that the ratio of the  $q^{th}$  amplitude to the first amplitude is  $r_q = |a_q| / |a_1|$ , then we have the following estimates for the component amplitudes:

$$|\hat{a}_q|^2 = \hat{r}_q^2 s_o / (1 + \sum_{i=2}^M \hat{r}_i^2) \quad (23)$$

where  $\hat{\phantom{x}}$  indicates the estimated values. Using the proposed TFD we can estimate the ratio  $r_q$  directly by eq.(3) at the peaks around the  $q^{th}$  and the first components,  $P_q(t, f)$  and  $P_1(t, f)$ , as follows

$$\hat{r}_q^2 = \text{mean}_{t,f} \{ | \frac{P_q(t, f)}{\psi'_q(f)} | \} / \text{mean}_{t,f} \{ | \frac{P_1(t, f)}{\psi'_1(f)} | \} \quad (24)$$

where  $\psi'_q(f)$  and  $\psi'_1(f)$  can be estimated after using the peak trajectory to estimate  $\phi'_q(t)$  and  $\phi'_1(t)$ , respectively. Since in discrete implementation the TFD builds up in the beginning and decays in the end due to the lack of correlation information, it is better not to include the start and the end parts of the TFD in the estimation using eq.(24). Also the regions of rapid change in the IF law should be excluded as eq.(3) would not be an accurate approximation to  $d(t, f)$  there (unless lag windowing is used). The best estimate is obtained when there is a linear part in the IF law, in this case the mean in eq.(24) is taken over this linear part, using eq.(4). Further studies on this amplitude estimate would be attempted in future works.

Using  $|\hat{a}_q|^2$  and  $\hat{\sigma}^2$  to calculate  $\text{var}(\Delta \hat{f}_{i,h_r}(t))$  (given by eq.(14) for  $d(t, f)$ ), we can define the confidence intervals  $\{D_{r,q}\}$  for all components as in [6]. The IF  $f_{i,q}(t)$  is contained in at least one of the confidence intervals  $\{D_{r,q}\}$  if  $h_r$  is sufficiently small, with a Gaussian probability  $P(\kappa)$ .

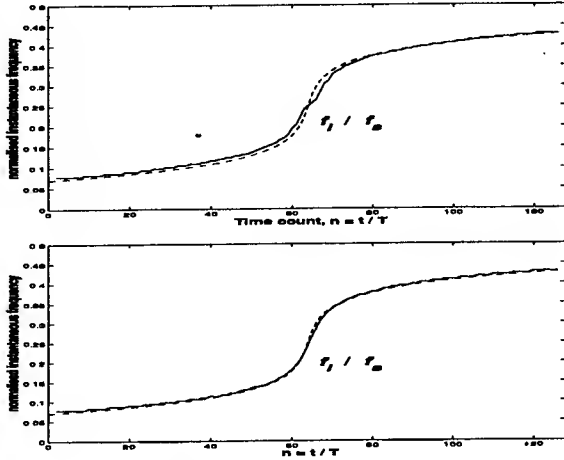


Figure 2: IF estimation of a mono-component non-linear FM signal with total signal length  $N = 128$  and  $T = 1/128$  using  $d(t, f)$ . Above: IF estimation using a constant window length  $h = 128$ . Below: Adaptive IF estimation. The estimated IF law is compared to the exact IF law (dashed line).

## 5. SIMULATION RESULTS

**Example 1:** The discrete version of  $d(t, f)$  as in eq.(9) using the time-lag kernel in eq.(1) is used to implement the adaptive algorithm for mono- and multi-component signals. For the mono-component case we consider a non-linear frequency-modulated signal with IF given by

$$f(nT) = 32 + 5 \sinh^{-1}(100(nT - 0.5))$$

with  $a = 1$ ,  $SNR = 15$  dB,  $\alpha = 0.1$ ,  $\kappa = 2$ ,  $0 \leq nT \leq 1$ , and  $T = 1/128$ . In Figure 2 The result of the adaptive IF estimation is shown as compared to the IF estimation using a constant window length. It can be noticed that the non-adaptive approach cannot estimate the IF law accurately at the instants of rapid change since the second derivative of the IF law is effective and the optimal window length as in eq.(16) is needed.

**Example 2:** We consider a two-component signal with non-linear frequencies given by

$$f_1(nT) = 40 + 5 \sinh^{-1}(20(nT - 0.4)), \text{ and :}$$

$$f_2(nT) = 20 + 2.5 \sinh^{-1}(50(nT - 0.8))$$

with  $a_1 = a_2 = 1$ ,  $SNR = 15$  dB,  $\alpha = 0.1$ ,  $\kappa = 2$ ,  $0 \leq nT \leq 1$ , and  $T = 1/128$ . In Figure 3 the result of the adaptive tracking algorithm is shown along with the adaptive window length for the first component. It is apparent that the adaptive window preserves lower lengths at the instants of rapid change in the component IF law in accordance with eq.(16).

## 6. CONCLUSIONS

This paper has presented an adaptive method to estimate the IF law of mono- and multi-component FM signals using quadratic time-frequency distributions. We proved that

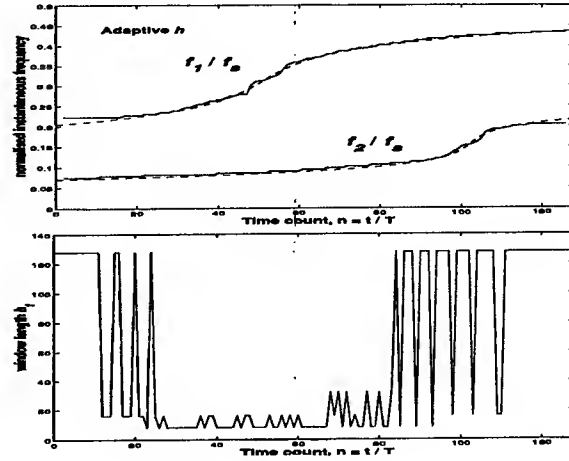


Figure 3: Above: adaptive IF estimation of a two-component FM signal using  $d(t, f)$  as compared to the exact IF laws (dashed lines). Below: adaptive window length as a function of time for the first component.

an IF estimation algorithm with adaptive window length is applicable to any quadratic time-frequency distribution that satisfies certain conditions. A time-frequency distribution  $d(t, f)$  that satisfies these conditions and enables non-parametric amplitude estimation is proposed. A comparison with a constant-window tracking algorithm shows that using a constant window length cannot give a robust IF estimate if the IF changes rapidly with time. A suggestion to adopt time-only kernels for the purpose of adaptive IF estimation is also presented.

## 7. REFERENCES

- [1] B. Boashash (editor), *Time-Frequency Signal Analysis*, Longman Cheshire, Melbourne, Australia, 1992.
- [2] B. Boashash, "Estimating and interpreting the instantaneous frequency of a signal," *Proc. IEEE*, vol. 8, no. 4, pp. 520-568, 1992.
- [3] L.J. Stankovic and V. Katkovnik, "Algorithm for the instantaneous frequency estimation using time-frequency distributions with adaptive window length," *IEEE Signal Processing Lett.*, vol. 5, no. 9, Sept. 1998.
- [4] V. Katkovnik and L. J. Stankovic, "Instantaneous frequency estimation using the Wigner distribution with varying and data driven window length," *IEEE Trans. Signal Processing*, vol. 46, no. 9, pp.2315-2325, 1998.
- [5] V. Sucic, B. Barkat, and B. Boashash, "Performance evaluation of the B-distribution," *Proceedings of the Fifth International Symposium on Signal Processing and Its Applications (ISSPA'99)*, Brisbane, Australia, vol. 1, pp. 267-270, Aug. 1999.
- [6] Z. M. Hussain and B. Boashash, "Adaptive instantaneous frequency estimation of multi-component FM signals," *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP'2000)*, Istanbul, June 5-9, 2000.



# DETECTION OF SEIZURES IN NEWBORNS USING TIME-FREQUENCY ANALYSIS OF EEG SIGNALS

Boualem Boashash, Helen Carson and Mostefa Mesbah

Signal Processing Research Centre (SPRC)  
Queensland University of Technology  
GPO Box 2434, Brisbane, Qld 4001, Australia

## ABSTRACT

This paper presents a time-frequency approach for electroencephalographic (EEG) seizure detection. The proposed method uses the high-resolution reduced interference B time-frequency distribution. An in-depth analysis of the seizure detection techniques of Gotman (frequency domain) and Liu (time domain) has been performed in order to compare with the detection criteria used in the time-frequency domain. Both synthetic and real neonatal EEG signals have been used for testing.

## 1. INTRODUCTION

Approximately one in every 200 newborn babies experiences some form of seizure, indicating cerebral abnormalities, or damage to the brain. Unlike adult seizure, the effects of newborn seizure are subtle and hence require the constant attention of a medical specialist for diagnosis.

Monitoring brain activity through electroencephalographic (EEG) data has become a successful means for detecting seizure in adults. This involves identifying sharp, repetitive waveforms in the EEG data that indicate the onset of seizure. In adults, these signals are easily recognisable against a low amplitude, random background characteristic of normal brain activity. The problem of detecting seizure in newborn babies, however, is complicated by many factors [1]. Firstly, healthy newborn EEG signals representing normal brain activity often contain patterns such as spurious waveforms and sharp spikes. These characteristics, which would otherwise be detected as seizure in adults, are simply the result of extra electrical activity produced by the immature brain as it continues to form. Seizures, however, still appear in the EEG data as repetitive waveforms and the problem lies in discerning the healthy spikes from those formed from seizures. Secondly, visual symptoms of seizure, such as muscle spasms, rapid eye movement, and drooling, are much more subtle in newborns and may be easily missed. These visual indicators are also natural movements common to all newborn babies. Thirdly, physical activity of babies in the intensive care environment is often subdued by medication to prevent injuries caused by unpredictable movements. This also eliminates the chance of seizure detection using visual signs altogether.

Currently there are three published methods for EEG seizure detection in newborns. The SPRC technique of Roessgen et al [1] is a parametric approach based on a nonlinear estimation of 11 model parameters for detection. The two other methods are non-parametric. The technique of Gotman [2] uses frequency analysis to determine the changes in the dominant peak of the frequency spectrum of short epochs of EEG data. The technique

of Liu [3] performs analysis in the time domain and is based on the auto-correlation function of short epochs of EEG data.

All three techniques are based on the assumption that the EEG signals are stationary or at least locally stationary. However, a closer examination of these signals often shows that EEG signals exhibit significant non-stationary and multi-component features (see figure 1).

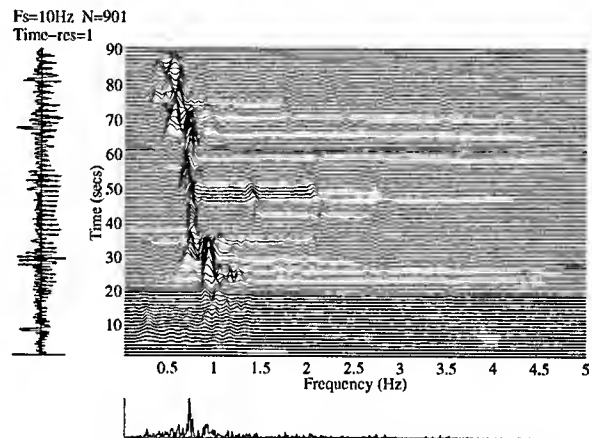


Figure 1 Time-frequency representations of newborn EEG seizure signal using the B-distribution

To take these characteristics into account, we propose in this paper a time-frequency (TF) domain approach. A prerequisite is the selection of an appropriate time-frequency distribution (TFD) that is capable of handling multi-component signals. Once the suitable TFD is chosen, a calibration process is undertaken. This involves initially reproducing the seizure detection criteria and characteristics used previously in other methods such as Gotman's and Liu's and map them in a joint time-frequency domain. Features in the  $t$ - $f$  domain indicating seizure are then identified and a detection process constructed and tested in the time-frequency domain. The proposed process is shown in figure 2.

## 2. DATA ACQUISITION

Electrical signals produced in the brain can be monitored in a non-invasive manner by measuring variations in potential on the scalp. This EEG measurement is achieved by strategically placing several small electrodes on the scalp, and forming a contact using conductive gel. One electrode, usually at the base of the skull, acts as a reference (ground) signal, and various channels of data are created by measuring the voltage differences between neighbouring electrodes.



Data used in our study has been collected at the Royal Women's Hospital Perinatal Intensive Care Unit in Brisbane, Australia\*. Due to the size of most newborn babies' heads, only five channels of EEG have been recorded in each session using the 10-20 International System of Electrode Placement. The EEG data has been recorded using a sampling frequency of 256 Hz. For artefact detection, three auxiliary signals representing electro-oculogram (EOG), electrocardiogram (ECG), and respiration are also recorded simultaneously with the EEG.

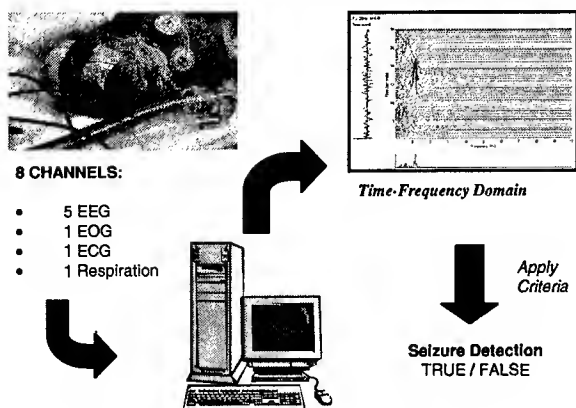


Figure 2 Time-frequency-based seizure detection process

### 3. TIME-FREQUENCY DISTRIBUTION SELECTION

In order to develop seizure detection methods in the time-frequency domain, it is necessary to select a suitable TFD to represent EEG data. Since neonatal EEG signals are non-stationary and occasionally multi-component, a desirable time-frequency distribution should have a good spectral resolution and reduced cross-terms. The performance and characteristics of several distributions have been compared to find an optimal representation of real neonatal EEG data in the time-frequency domain. The scope of this comparison study has encompassed seven distributions, including the Spectrogram, Wigner-Ville, Choi-Williams, B-Distribution, Zalto-Atlas-Marks, Born-Jordan, and Rihaczek-Margenau distributions [4-5]. Each time-frequency distribution has been applied to epochs of real neonatal EEG for various data window lengths and individual TFD parameter values. The performances of the resulting time-frequency representations have been compared using an objective quantitative measure criterion [5]. Based on this criterion, the B-distribution with the smoothing parameter equal to 0.01 has been selected as the most suitable representation of the EEG signals in the time-frequency domain [5]. Figure 3 illustrates the time-frequency representations of a 30-seconds sample of real newborn EEG data using the B and the Choi-Williams (CW) distributions.

### 4. FROM TIME DOMAIN TO TIME-FREQUENCY DOMAIN

#### 4.1 Review of Liu's Method

In his method, Liu relied on the assumption that the essential characteristic in newborn seizure EEG is periodicity. The amount

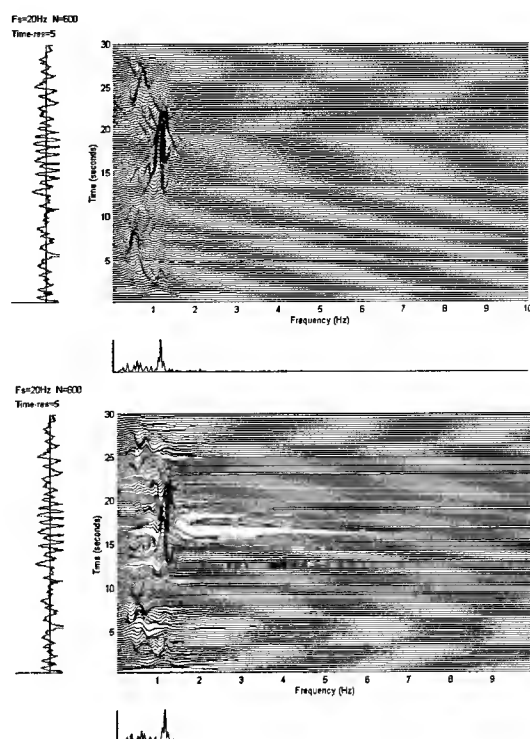


Figure 3 The B-distribution with  $\beta = 0.01$  (above) and the CW distribution with  $\sigma = 10$  (below) of a real epoch of newborn EEG.

of periodicity in the autocorrelation of short epochs of EEG data is scored and used in a rule based algorithm to perform classification. In this technique, an epoch consisting of 30 seconds of data is divided into 5 windows (see figure 4).

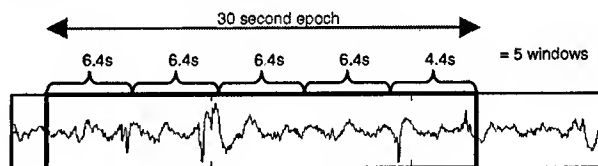


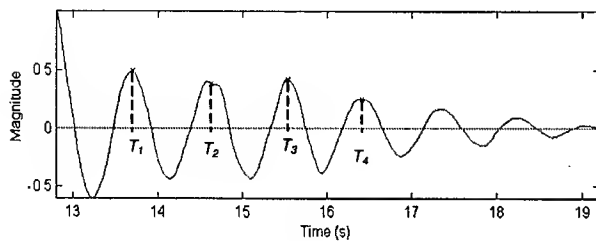
Figure 4 Epoch and window definitions according to Liu

Depending on the autocorrelation function of each window, up to four primary periods ( $T_1, \dots, T_4$ ) are calculated for each window in an epoch as shown in figure 5. These times correspond to the moment centres of the first, second, third and fourth peaks in the autocorrelation function. The windows are then scored whereby more evenly spaced primary periods are allocated larger scores. After each window in an epoch is scored, a rule based detection scheme is applied to classify each epoch as positive or negative. If two or more channels of EEG data in the same epoch are as positive, the epoch is then classified as containing seizure.

The above procedure can be summarised as follows:

- Calculate the autocorrelation function for each window within each epoch.

- Locate the first four moment centres between zero-crossings (if they exist).
- Calculate the ratios between the first and subsequent moment locations.
- Find the differences between each of these three ratios and the nearest integer.
- Assign scores to each difference according to a non-linear scoring system
- Sum these three scores to give a total for each window in each epoch
- Label as seizure positive or negative, depending upon their scores, the different windows within each epoch and across all channels



**Figure 5** Autocorrelation function for one window

## 4.2 Seizure criteria in time domain

A single window is seizure positive by the method of Liu if the following criteria are met:

- At least four periods exist within the positive half of the Autocorrelation function.
- The differences between the ratios of each moment centre to the first and the nearest integer are less than 0.150.
- The total score obtained by summing all moment centre scores is greater than or equal to 12 (out of a maximum of 15).

These criteria have been determined after closely analysing the scoring system to find definite constraints defining seizure from non-seizure signals in the time domain. This has essentially been achieved by firstly identifying all scoring scenarios leading to a score greater or equal to +12, the simplest way of achieving seizure detection as defined by Liu. Secondly, these scores may be broken down into the forerunning ratios and moment centre calculations necessary to achieve each score. Finally, inherent signal characteristics and constraints necessary to achieve these moment centres in the lag domain can be identified.

## 4.3 Seizure criteria in time-frequency domain

An EEG signal within a given window is considered seizure if a continuous spectral peak exists within the window and meets the following criteria:

- All frequencies within the spectral line are greater than 0.625 Hz within 6.4-second windows or greater than 0.909 Hz within 4.4-second windows.
- The length of continuous dominant spectral peak within the window is greater than 3 seconds.

where a continuous spectral peak is defined as adjoining peaks within the time-frequency array above a threshold of one fifth the maximum array value.

The first criterion is a direct translation to the frequency domain of the first criterion described in section 4.2. That is, in order for three moment centres to exist, four periods must occur in the lag domain over a single window. Applying the property of the Autocorrelation function that states that the Autocorrelation function of a periodic signal is also periodic with the same period, this is interpreted by assuming four periods of the signal must exist in a single window. This translates simply into the first time-frequency criteria stated above.

The second criterion has been deduced by observation of several time-frequency representations of seizure positive windows as defined by Liu. The scoring system focuses on identifying periodic regions of data using the lag domain. Periodic regions are clearly identified in time-frequency representations by a dominant spectral peak occurring for a certain time interval. Therefore, this is a less stringent criteria translation, but one based upon identifying a common characteristic in each domain. Further statistical analysis is required to determine an exact duration necessary to identify seizure by this method. The key factor to the success of this method has involved the discovery of frequency restrictions existing inherently within the scoring system designed by Liu.

## 4.4 Implementation

Extraction of the seizure criteria listed above in the TF domain has been successfully calibrated for the method of Liu. Peak detection techniques from image processing have been employed to simplify the extraction process, resulting in a detection array illustrating positions and lengths of continuous spectral lines within each epoch. Figure 6 shows the algorithm flow chart used in the implementation of this method.

## 4.5 Results and discussion

Very promising results have been obtained using time-frequency algorithms to detect individual seizure windows of real neonatal EEG in the time-frequency domain. Approximately 75% of windows detected as seizure positive by Liu are detected by applying TF criteria listed above. The result of applying the proposed time-frequency-based detection method is illustrated in figure 7. Original epoch refers to the raw TF array produced from pre-processed EEG data. Images of these arrays appear on the left side of the figure. These arrays are also divided into four distinct 6.4-second windows and one 4.4-second window as defined by Liu. Window scores attained for these epochs are displayed at the end of each window division. This makes for an easy comparison between the TF information contained within each window, and the corresponding score allocated by Liu.

That only 75% of the seizures predicted by Liu's method are detected is mainly due to the fact that our method uses scores of single windows while Liu used the combined scores of up to three consecutive windows per epoch. Future implementations of our method will include the different possible window combinations.

## 5. FROM FREQUENCY DOMAIN TO TIME-FREQUENCY DOMAIN

### 5.1 Review of Gotman's method

The method proposed by Gotman is based on spectral analysis and is used to detect periodic discharges. A background epoch is defined as a 20-second segment of EEG finishing 60 seconds before the start of the current 10-second epoch being investigated (see figure 8). The main advantage of a moving background epoch is that results are not dependent on the specific features of a fixed epoch.

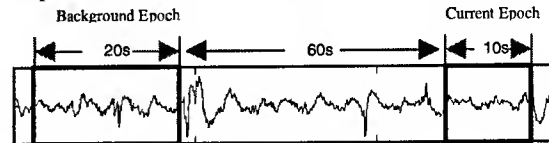


Figure 8 Epoch and window definitions according to Gotman

The frequency spectrum of each 10-second epoch is calculated and the following features are extracted:

- The frequency of the dominant spectral peak.
- The width of the dominant spectral peak.
- The ratio of the power in the dominant spectral peak to that of the background spectrum in the same frequency band.

The 10-second epoch of data is considered seizure positive if any of the following criteria are met:

	Dominant Frequency	Half-Maximum Bandwidth	Power Ratio
1.	0.5 – 1.5 Hz	$\leq 0.6$ Hz	3 – 4
2.	1.5 – 10 Hz	$\leq 0.6$ Hz	2 – 4
3.	1.5 – 10 Hz	$\leq 1$ Hz	4 – 80

If an epoch is classified as containing seizure based on the above criteria, a further three criteria are used to limit the number of false alarms. Seizure detection is discounted if the epoch is largely non-stationary, if there is a large amount of AC power noise present or if it appears that an EEG lead has been disconnected.

The aim of this method is to determine if a dominant peak exists in the power spectral density (PSD). This is equivalent to detecting if an EEG waveform has a dominant periodic shape in the time domain. The feature space used to classify an epoch as seizure ensures that the dominant peak of the spectrum is significant compared to the background spectrum.

### 5.2 Seizure criteria in time-frequency

Since a time-frequency representation is comprised of the instantaneous spectra of a signal over time, criteria pertaining to frequency and bandwidth above are clearly discernible in the time-frequency array. That is, each spectra containing a dominant peak that meets either of the criteria:

- Frequency in the range 0.5 – 1.5 Hz and width  $\leq 0.6$  Hz.
- Frequency in the range 1.5 – 10 Hz and width  $\leq 1$  Hz.

may be considered for further seizure detection pertaining to power ratio. Disregarding the power ratio criteria defined in section 5.1, the second criterion becomes a subset of third criterion. Due to the instantaneous nature of the time-

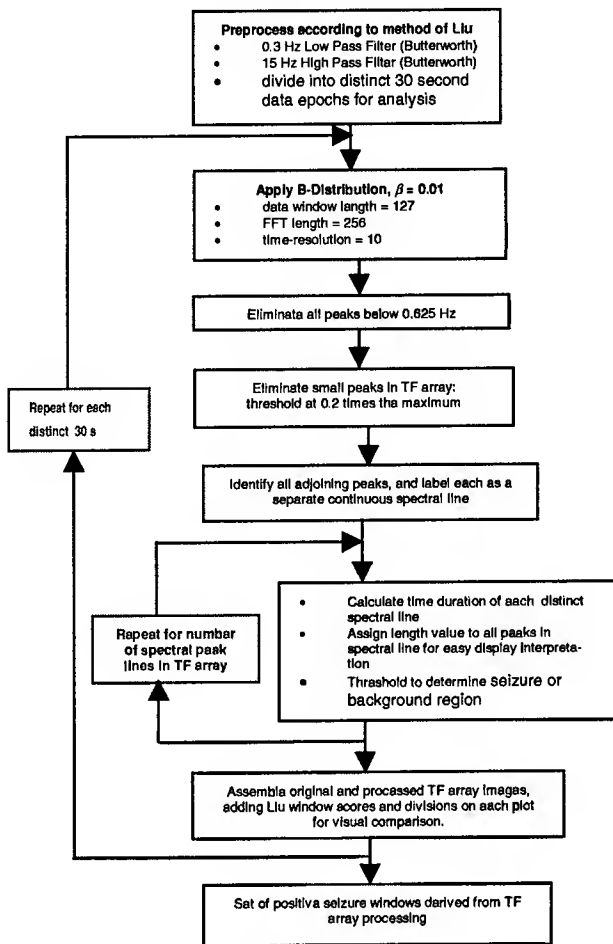


Figure 6 Implementation and calibration of the time-frequency extension of time-domain method

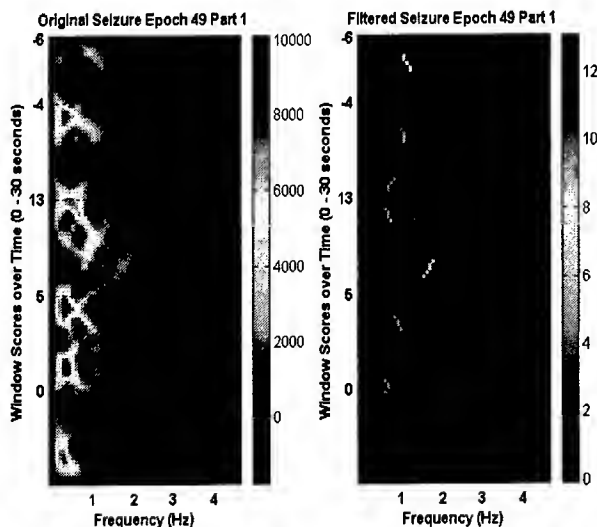


Figure 7 The mapping of Liu's method to time-frequency

frequency array, this power ratio is not obvious, and requires further investigation for explicit definition.

### 5.3 Implementation

Algorithm implementing the two time-frequency criteria identified above is illustrated in figure 9. This essentially extracts frequency and width information, the results of which are visible in the plots shown in figure 10.

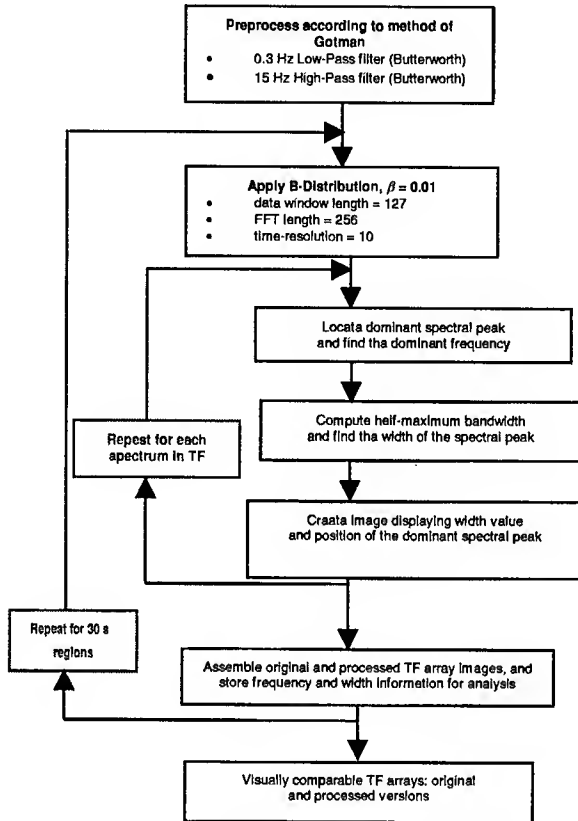


Figure 9 Implementation and calibration of time-frequency extension of frequency-based method

### 5.4 Results and discussion

The result of applying the above proposed algorithm is illustrated in figure 10. Data is presented by highlighting the position of the dominant frequency with a colour indicating the width of the spectral peak. Boxed sections of the array indicate regions detected as containing seizure by the conventional method of Gotman. This has been included to aid visual recognition of any predominant features that may stand out in the processed time-frequency array of seizure epochs. The main limitations of this method lie in the ability to accurately assess differences in power between current and reference epochs due to the instantaneous nature of the time-frequency array under analysis. Further research into this matter, and its incorporation into the detection algorithm defined in the above sections, should result in clearly recognisable features defining seizures in the TF array.

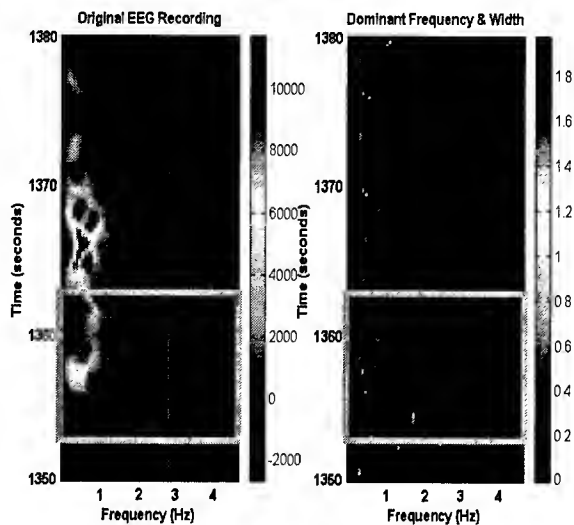


Figure 10 The mapping of Gotman's method to time-frequency

## 6. CONCLUSIONS

The initial results obtained show that the time-frequency domain is a suitable basis from which to develop a complete seizure detection scheme. Successful mapping of Liu's detection criteria into the time-frequency domain has been completed, and mapping of two out of three criteria detailed by Gotman have also been implemented and tested in the time-frequency domain. Those two mappings allowed us to calibrate the time-frequency-based method. The next step will be to develop a fully integrated time-frequency detection method by combining the different time-frequency seizure features identified in this paper.

Essentially, this paper provides proof of concept of seizure detection in the time-frequency domain. Further results will appear elsewhere.

## 7. REFERENCES

- [1] M. Roessgen, A. Zoubir, B. Boashash. "Seizure Detection of Newborn EEG Using a Model Based Approach", *IEEE Trans. on Biomed. Eng.*, Vol 45, No 6, June 1998.
- [2] J. Gotman, J. Zhang, J. Rosenblatt, R. Gottsman. "Evaluation of an Automatic Seizure Detection Method for the Newborn EEG", *Electroenc. and Clin. Neurophysy.*, 103:363-369, 1997.
- [3] A. Liu, J. S. Hahn, G. P. Heldt, R. W. Coen. "Detection of Neonatal Seizures through Computerized EEG Analysis". *Electroenc. and Clin. Neurophysy.*, 82:30-37, 1992.
- [4] B. Boashash, "Time-Frequency Signal Analysis", in S. Haykin, editor, *Advances in Spectral Estimation and Array Processing*, Prentice Hall, 418-517, 1991.
- [5] B. Boashash and V. Sucic. "A Resolution Performance Measure for Quadratic Time-Frequency Distributions", *Proc. of 10<sup>th</sup> IEEE Stat. Sig. and Array Proc.*, 2000.

\* The authors acknowledge the contributions of Dr. Paul Colditz in providing access to the Perinatal Research Centre and the interpretation of the EEG data, Mr. Mark Keir for data collection and Dr. Patrick Celka for pre-processing the signal in figure 1 and for reviewing the paper.

# MULTITAPER REDUCED INTERFERENCE DISTRIBUTION

*Selin Aviyente and William J. Williams*

Department of Electrical Engineering and Computer Science,  
The University of Michigan, Ann Arbor, MI 48109-2122  
e-mail: saviyent@eecs.umich.edu

## ABSTRACT

Time-frequency distributions (TFDs) belonging to Cohen's class are usually designed to satisfy the frequency marginal. The conventional frequency marginal is equivalent to the classical periodogram. It is known that the periodogram is not a good spectral estimator. For this reason, Thomson [3] introduced a multitaper spectral estimator which treats both the bias and the variance problems inherent to nonparametric spectral estimators. In this paper, we are introducing a new kernel design method which achieves Thomson's spectral estimate as the frequency marginal. This new method results in a signal-dependent kernel design. The resulting kernel belongs to the class of reduced interference distribution (RID) kernels and therefore this new time-frequency distribution will be called multitaper reduced interference distribution (MT-RID). The performance of this method is compared to the previously introduced multiwindow time-frequency distribution (MW-TFD) [6] through simulations.

## 1. INTRODUCTION

Time-frequency distributions belonging to Cohen's class [1] are usually designed to satisfy the frequency marginal. The time-frequency kernel is designed such that it yields the classical periodogram as the frequency marginal. It is known that the classical periodogram is not a good spectral estimator due to its nonzero variance even when the number of data samples goes to infinity [2]. For this reason, several modified periodogram methods have been introduced which reduce the variance at the expense of increasing its bias. For short, time-limited signals, Thomson [3] suggested using a set of orthogonal windows to compute several direct spectrum estimates of the entire signal and then average the resulting spectrums to construct a spectral estimate. The orthogonal windows used are the eigenfunctions, discrete prolate spheroidal wave functions [4], of the spectral estimation kernel. Since the windows are orthogonal and optimally concentrated in the frequency domain, the resulting spectral estimate treats both the bias and the variance problems.

Several authors have extended Thomson's method to nonstationary spectrum estimation [5, 6, 7, 8]. In [6] and [7], the authors applied prolate spheroidal sequences

or windows which are optimally concentrated in the time-frequency plane, i.e. Hermite functions, to compute several spectrograms and then combined them to obtain a time-varying spectrum estimate. In [5], the spectral representation theorem for stationary processes is extended to the nonstationary case and the eigenfunctions are found to construct the time-varying spectrum estimator.

In this paper, we approach the problem from the perspective of the frequency marginal and solve for a time-frequency kernel which will give us the desired marginal. We derive the conditions on the time-frequency kernel such that it yields Thomson's spectrum as the frequency marginal. It is shown that the corresponding time-frequency kernels are signal-dependent. The time-frequency kernels designed in this manner belong to the class of reduced interference distribution (RID) kernels. Therefore, we are going to refer to this new class of time-frequency distributions as Multitaper-RID (MT-RID). This approach provides smoother time-frequency distributions which are less prone to noise. The performance of this method is then compared to the multiple window spectrogram method [6] for example signals in additive noise through simulations.

## 2. MULTITAPER REDUCED INTERFERENCE DISTRIBUTION (MT-RID)

For a time-frequency distribution belonging to Cohen's class, it's desirable to have the following two properties, the time and the frequency marginal, satisfied<sup>1</sup>:

$$\begin{aligned}\int C(t, \omega) d\omega &= |s(t)|^2 \\ \int C(t, \omega) dt &= |S(\omega)|^2\end{aligned}\quad (1)$$

When the frequency marginal is satisfied, the energy distribution of the signal is represented by the classical periodogram  $|S(\omega)|^2$ . It is well known that the classical periodogram is not a good spectral estimator due to its inconsistency. Thomson's method overcomes the bias-variance tradeoff inherent in nonparametric spectral estimation methods. This method is equivalent to using the weighted average of a series of direct spectrum estimates based on orthogonal data windows. The high resolution spectrum estimate around a point  $f_0$  is

This research was supported in part by grants from the Rackham School of Graduate Studies and the Office of Naval Research, ONR grant no. N000014-97-1-0072

<sup>1</sup>All integrals are from  $-\infty$  to  $\infty$  unless otherwise specified.

$$\overline{S(f_o)} = \frac{1}{2NW} \sum_{k=0}^{K-1} \frac{1}{\lambda_k(N, W)} |y_k(f_o)|^2 \quad (2)$$

where  $\lambda_k$  is the eigenvalue corresponding to the  $k$ th orthogonal window,  $NW$  is the time-bandwidth product and each  $|y_k(f_o)|^2$  is the spectral estimate computed using the  $k$ th discrete prolate spheroidal sequence expressed as:

$$|y_k(f_o)|^2 = \left| \sum_{n=0}^{N-1} x(n) \frac{v_n^{(k)}(N, W)}{\epsilon_k} e^{-j2\pi f_o n} \right|^2 \quad (3)$$

where  $v_n^{(k)}(N, W)$ s are the discrete prolate spheroidal sequences and  $\epsilon_k$  is a normalization factor.

We take this spectrum estimate as a basis for our improved frequency marginals, and solve for the corresponding time-frequency kernel.

Any time-frequency distribution with an alias-free kernel [10],  $\psi(n, m)$  in the time-lag domain can be written as <sup>2</sup>:

$$\sum_m \sum_l \psi(n-l, m) x(l + \frac{m}{2}) x^*(l - \frac{m}{2}) e^{-j\omega m} \quad (4)$$

First, let us equate the frequency marginal of this distribution to a smoothed periodogram, i.e. data is smoothed by a single window, and then extend the results to Thomson's multitaper spectrum estimate. The frequency marginal is equal to

$$\begin{aligned} \sum_n \sum_m \sum_l \psi(n-l, m) x(l + \frac{m}{2}) x^*(l - \frac{m}{2}) e^{-j\omega m} \\ = \left| \sum_k h(k) x(k) e^{-j\omega k} \right|^2 \end{aligned} \quad (5)$$

where  $h(k)$  is the time domain window. After some algebraic manipulations, it is found that

$$\begin{aligned} \sum_m \sum_k h(k + \frac{m}{2}) x(k + \frac{m}{2}) h^*(k - \frac{m}{2}) x^*(k - \frac{m}{2}) e^{-j\omega m} \\ = \sum_m \sum_n \sum_l \psi(n-l, m) x^*(l - \frac{m}{2}) x(l + \frac{m}{2}) e^{-j\omega m} \end{aligned} \quad (6)$$

This equality implies that

$$\begin{aligned} \sum_l h(l + \frac{m}{2}) x(l + \frac{m}{2}) h^*(l - \frac{m}{2}) x^*(l - \frac{m}{2}) \\ = \sum_l \left[ \sum_n \psi(n-l, m) \right] x^*(l - \frac{m}{2}) x(l + \frac{m}{2}) \quad \forall m \end{aligned} \quad (7)$$

If we express the kernel in terms of its ambiguity domain function, we will end up with the following expression:

$$\begin{aligned} \sum_l \left[ \sum_n \int \phi(\theta, m) e^{j\theta(n-l)} d\theta \right] x(l + \frac{m}{2}) x^*(l - \frac{m}{2}) \\ = \sum_l \left[ \int \left( \sum_n e^{j\theta n} \right) \phi(\theta, m) e^{-j\theta l} d\theta \right] x(l + \frac{m}{2}) x^*(l - \frac{m}{2}) \end{aligned} \quad (8)$$

<sup>2</sup>All summations are  $\sum_{-\infty}^{\infty}$  unless otherwise specified.

If we further assume that the number of signal samples goes to infinity then the above equation will reduce to:

$$\begin{aligned} \sum_l \phi(0, m) x(l + \frac{m}{2}) x^*(l - \frac{m}{2}) \\ = \sum_l h(l + \frac{m}{2}) x(l + \frac{m}{2}) h^*(l - \frac{m}{2}) x^*(l - \frac{m}{2}) \quad \forall m \end{aligned} \quad (9)$$

which will in turn give an expression for  $\phi(0, m)$  in terms of the signal and the data window used in spectral estimation.

$$\phi(0, m) = \frac{\sum_l h(l + \frac{m}{2}) x(l + \frac{m}{2}) h^*(l - \frac{m}{2}) x^*(l - \frac{m}{2})}{\sum_l x(l + \frac{m}{2}) x^*(l - \frac{m}{2})} \quad (10)$$

The above equation can be interpreted to be the ratio between the biased autocorrelation estimate for the windowed signal and the biased autocorrelation estimate for the original signal. It is important to keep in mind that the above equation is only an asymptotic result, since it is only true when the number of signal samples goes to infinity. Since in real cases we are going to have a finite number of samples of a given signal, the equality given above will only be an estimate of the actual result. This result also implies that when we have a rectangular window for  $h(t)$ ,  $\phi(0, m)$  becomes equal to 1 which is the well-known constraint on time-frequency kernels for satisfying the conventional frequency marginal. For anything other than the rectangular window, it is not possible to have general constraints on the kernel. Therefore, the kernel designed to satisfy a specific frequency marginal will be signal dependent.

We can easily extend these results to Thomson's spectrum estimate by combining the constraints imposed by each window.

If we apply the previous results for the time-frequency kernel to achieve a frequency marginal equal to  $|y_k(f_o)|^2$  given in equation 3, we will obtain:

$$\phi_k(0, m) = \frac{\sum_l v^{(k)}(l + \frac{m}{2}) v^{*(k)}(l - \frac{m}{2}) x(l + \frac{m}{2}) x^*(l - \frac{m}{2})}{\sum_l x(l + \frac{m}{2}) x^*(l - \frac{m}{2})} \quad (11)$$

where each data window produces its own corresponding kernel. The final kernel is a weighted summation of individual kernels.

$$\phi(0, m) = \frac{1}{2NW} \sum_{k=0}^{K-1} \frac{1}{\lambda_k(N, W)} \phi_k(0, m) \quad (12)$$

It is apparent from this equation that there is no unique way of designing the kernel given this one dimensional constraint. For this reason, we consider a construction algorithm which will require the least amount of computation. The kernel is built in an iterative manner in the time-lag domain such that the summation of the kernel elements along the time direction at any given time-lag gives the value of  $\phi(0, m)$  at the particular lag value [9]. This construction guarantees that the desired frequency marginal is achieved along with RID characteristics and provides minimal computational complexity since the kernel is constructed as an outer product of orthogonal vectors.

### 3. REVIEW OF MULTIWINDOW TIME-FREQUENCY DISTRIBUTION (MW-TFD)

Thomson's multiwindow spectrum estimation for stationary signals is extended to the nonstationary case in [6]. The MW-TFD is applied to a signal in a similar manner as the spectrogram. However, instead of applying a single sliding window along the signal, the MW-TFD applies a set of orthogonal sliding windows and then takes the average as follows.

$$X_{MW}(n, \omega) = \frac{1}{K} \sum_{k=0}^{K-1} |X_k(n, \omega)|^2 \quad (13)$$

where each  $X_k(n, \omega)$  is expressed as a short-time Fourier transform computed using the  $k$ th window function.

$$X_k(n, \omega) = \sum_m x(m) h_k(m-n) e^{-j2\omega m} \quad (14)$$

### 4. EXAMPLES

In this section, we are going to give some preliminary results for the bias and the variance analysis of the two time-varying spectral estimators discussed above. The main performance analysis will be based on simulations.

For the MW-TFD, the expected value of the estimate is given by:

$$E[X_{MW}(n, \omega)] = \frac{1}{K} \sum_{k=0}^{K-1} \sum_l \sum_m r(l, m) h_k(l-n) h_k^*(m-n) e^{-j\omega(l-m)} \quad (15)$$

When the windows are normalized, this results in an unbiased estimator for stationary white process. Similarly, for MT-RID, the expected value of the distribution is given as follows:

$$E[X_{MT-RID}(n, \omega)] = \sum_m \sum_l \psi(n - \frac{l+m}{2}, l-m) r(l, m) e^{-j\omega(l-m)} \quad (16)$$

When the kernel is properly normalized, we get an unbiased estimator for white spectra. At this point, we don't have closed form expressions for the variances of the two time-frequency distributions. Amin has introduced formulations for average variance of time-frequency distributions of signals in noise in [11]. Since we are interested in analyzing local phenomena, the variance of the time-frequency distributions will be compared through simulations. The first example that we will consider is a complex exponential with additive white Gaussian complex noise. The signal plus noise model can be expressed as:

$$\begin{aligned} x(n) &= s(n) + \eta(n) \quad n = 0, \dots, 64 \\ s(n) &= \exp(j\omega_0 n) \\ \text{Var}[\eta(n)] &= 0.1 \end{aligned} \quad (17)$$

In this case, the kernel designed to achieve Thomson's spectrum as the frequency marginal, equation 12, has to satisfy the following condition:

$$\phi_k(0, m) = \sum_l h_k(l + m/2) h_k^*(l - m/2)$$

$$\phi(0, m) = \frac{1}{K} \sum_{k=0}^{K-1} \phi_k(0, m) \quad (18)$$

This is equivalent to averaging the autocorrelation functions of individual windows and imposing that as the kernel at  $\theta = 0$ . Similarly, the kernel for MW-TFD is the average of the ambiguity functions,  $A_{h_k}(\theta, m)$ , for each window. It is expressed as:

$$\begin{aligned} \phi_{MW}(\theta, m) &= \frac{1}{K} \sum_{k=0}^{K-1} A_{h_k}(-\theta, m) \\ A_{h_k}(-\theta, m) &= \sum_l h_k(l + m/2) h_k^*(l - m/2) e^{-j\theta l} \end{aligned} \quad (19)$$

It is apparent from the above two equations that these two kernels agree for  $\theta = 0$ , and thus will have similar frequency marginals. The kernel for MW-TFD in the ambiguity domain is concentrated along  $\theta = 0$  axis, whereas the MT-RID kernel will have RID structure due to the design procedure described in Section 2. (Figure 1) The structure of the kernel for MW-TFD suggests that it is good in extracting impulses along the frequency dimension, i.e. complex exponentials, and not so good in tracking time-varying phenomena.

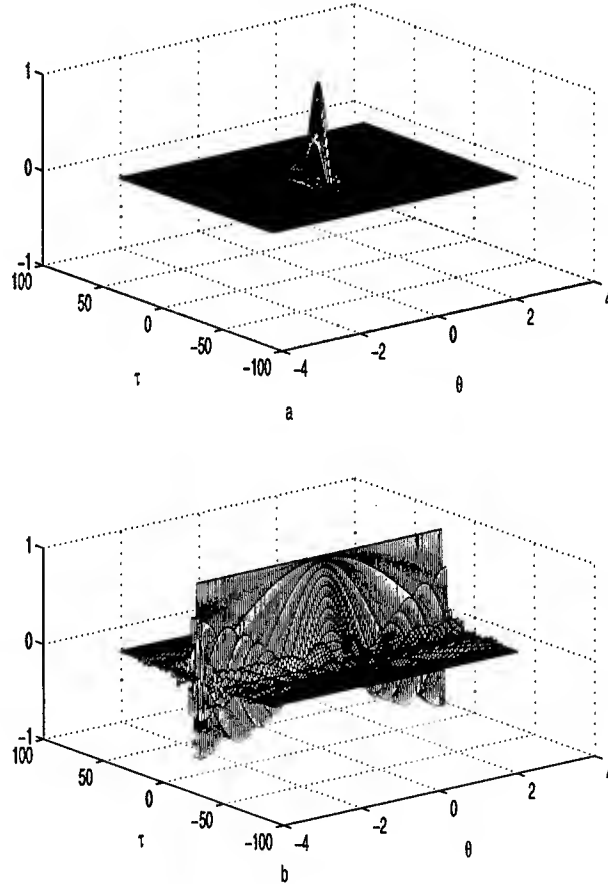


Figure 1: The time-frequency kernels in the ambiguity domain for example 1 a) The kernel for MW-TFD, b) The kernel for MT-RID

The complex exponential signal buried in white Gaussian complex noise is simulated at SNR=10dB for 500 times. The standard deviation is computed across all time for all simulations. It is seen that the deviation is highest at the fundamental frequency of the complex exponential. (Figure 2) MT-RID produces a time-frequency distribution with a higher variance compared to the MW-TFD since the latter is a spectrogram-based method which only takes on positive values. In the second example, we consider a linear chirp

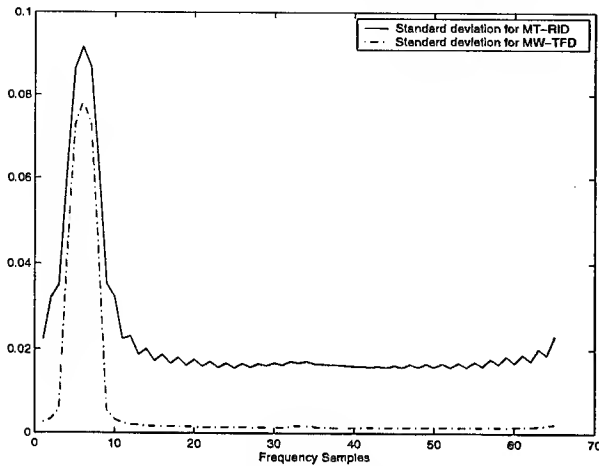


Figure 2: The standard deviations of the time-frequency distributions for the complex exponential signal plus noise based on 500 simulations.

signal plus noise.

$$\begin{aligned} x(n) &= s(n) + \eta(n) \quad n = 0, \dots, 64 \\ s(n) &= \exp(j(\omega_0 n + \beta n^2/2)) \\ \text{Var}[\eta(n)] &= 0.1 \end{aligned} \quad (20)$$

In this case, the kernels for the two methods are not the same and a closed form expression for the constraint on the kernel of MT-RID is complicated to formulate. Still, we can build the kernel using equation 12. An analysis similar to above can be done to see the standard deviation of the time-frequency distribution based on 500 simulations. In this case, the MT-RID method exhibits a more stable spectrum. Due to the shape of its kernel in the ambiguity domain, the MW-TFD method does not offer good resolution properties. This induces a large variance around the true chirp rate. (Figure 3)

## 5. CONCLUSIONS

In this paper, we have introduced an alternative approach to extending Thomson's multitaper spectrum estimation method to the time-varying case. The necessary condition on the kernel function to obtain a frequency marginal equal to Thomson's spectrum estimator is derived and this leads to a new time-frequency analysis method, multitaper reduced interference distribution (MT-RID). This method is then compared to the MW-TFD method which is a direct extension of Thomson's method to nonstationary case [6]. The statistical performance of the two methods are compared for noisy test signals through simulations. The

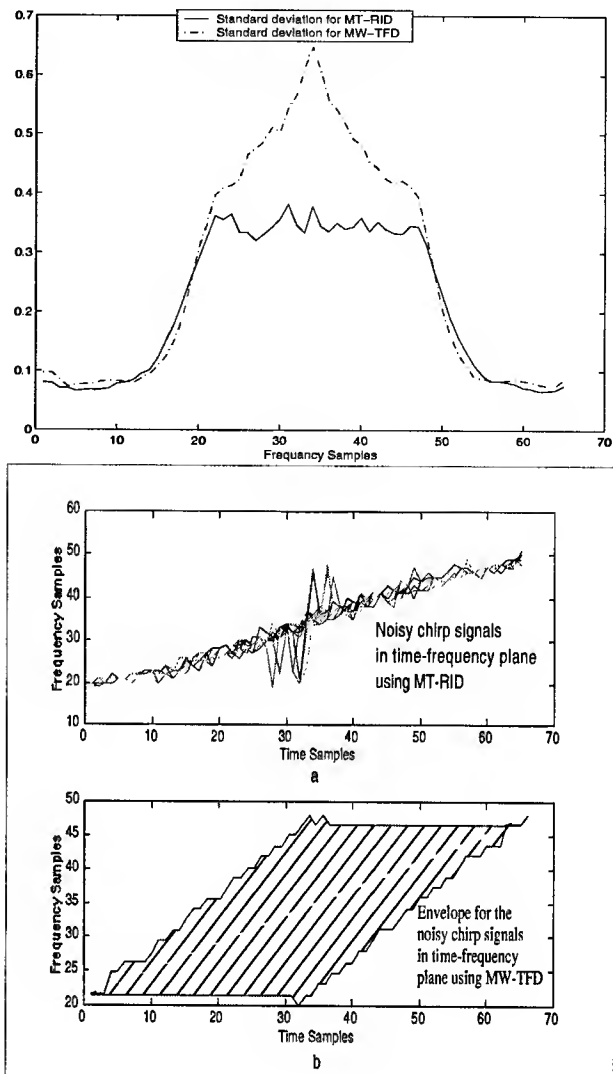


Figure 3: The standard deviation of the frequency marginals for the noisy chirp signal and the variance of time-frequency distributions for 500 simulations: a) MT-RID method, b) MW-TFD method

results show that MT-RID gives a better resolution for time-varying components whereas the MW-TFD is better for monochromatic signals. The MW-TFD also offers a smoother distribution due to extensive averaging inherent to its mechanism. The results can be generalized for different classes of signals by obtaining an expression for the variance of the estimators.

## 6. REFERENCES

- [1] L. Cohen, *Time-Frequency Analysis*. Prentice Hall, 1995.
- [2] P. Stoica and R. Moses, *Introduction to Spectral Analysis*. Prentice Hall, 1997.
- [3] D. J. Thomson, "Spectrum Estimation and Harmonic Analysis," *Proc. of the IEEE*, vol. 70, no. 9, pp.1055-1096, September 1982.



- [4] D. Slepian, "Prolate spheroidal wave functions, Fourier analysis, and uncertainty-V: The discrete case," *Bell Syst. Tech. J.*, vol.57, pp.1371-1430, 1978.
- [5] J. W. Pitton, "Time-Frequency Spectrum Estimation: An Adaptive Multitaper Method," in *Proc. IEEE-SP Int. Sym. Time-Frequency and Time-Scale Analysis*, pp.665-668, October 1998.
- [6] G. Frazer and B. Boashash, "Multiple window spectrogram and time-frequency distributions," in *Proc. IEEE Int. Conf. Acoustic, Speech, and Signal Processing-ICASSP'94*, vol.IV, pp. 293-296, 1994.
- [7] F. Çakrak and P. Loughlin, "Multiple window non-linear time-varying spectral analysis," in *Proc. IEEE Int. Conf. Acoustic, Speech, and Signal Processing-ICASSP'98*, vol.IV, pp. 2409-2412, 1998.
- [8] M. Bayram and R. G. Baraniuk, "Multiple window time-frequency analysis," in *Proc. IEEE-SP Int. Sym. Time-Frequency and Time-Scale Analysis*, pp.173-176, 1996.
- [9] W. J. Williams and S. Aviyente, "Optimal Window Time-Frequency Distribution Decompositions," 32nd Asilomar Conference on Signals, Systems and Computers, pp.817-821, 1998.
- [10] J. C. O'Neill, P. Flandrin and W. J. Williams, "On the Existence of Discrete Wigner Distributions," *IEEE Signal Processing Letters*, vol. 6, no. 12, pp.304-306, December 1999.
- [11] M.G. Amin, "Minimum Variance Time-Frequency Distribution Kernels for Signals in Additive Noise," *IEEE Trans. Signal Processing*, vol.44, no. 9, pp.2352-2356, September 1996.

# INSTANTANEOUS SPECTRAL SKEW AND KURTOSIS\*

*Patrick J. Loughlin and Keith L. Davidson*

Dept. of Electrical Engineering, 348 Benedum Hall  
University of Pittsburgh, Pittsburgh, PA 15261  
E-mail: pat@ee.pitt.edu  
Tel: (412) 624-9685 Fax: (412) 624-8003

## ABSTRACT

We extend the ideas of the instantaneous frequency and the instantaneous bandwidth of a signal by defining the instantaneous skew and kurtosis, as well as higher instantaneous moments, of a signal. Expressions are derived in terms of the signal amplitude and phase, analogous to the situation for instantaneous frequency and bandwidth. As is the case in time-frequency analysis with instantaneous frequency and bandwidth, the instantaneous moments we derive may be viewed as conditional moments of the time-varying spectral density of the signal.

## 1. INTRODUCTION

Instantaneous frequency is a fundamental concept of signals arising in many areas, including communications, seismology, sonar and radar, among others [1, 2, 7, 14, 18]. As Ville showed, it is intimately connected to the time-varying spectrum of a signal; in particular, Ville showed that the first conditional spectral moment of the Wigner distribution of a signal  $s(t) = A(t)e^{j\varphi(t)}$  is equal to its instantaneous frequency,  $\varphi'(t)$  [18]. This property was later found to hold for many other time-frequency distributions of a signal [4, 7].

The conditional spectral moments of a time-frequency distribution  $P(t, \omega)$  are obtained, as with any joint density, via

$$\begin{aligned}\langle \omega^n \rangle_t &= \frac{1}{P(t)} \int \omega^n P(t, \omega) d\omega \\ &= \int \omega^n P(t, \omega) d\omega / \int P(t, \omega) d\omega.\end{aligned}\quad (1)$$

Central conditional moments can also be obtained in the usual way, by subtracting the mean frequency at each time, according to

$$\mu_\omega^n(t) = \frac{1}{P(t)} \int (\omega - \langle \omega \rangle_t)^n P(t, \omega) d\omega. \quad (2)$$

\*This work was supported by the Office of Naval Research (grant no. N00014-98-1-0680).

Cohen has extensively considered the second central conditional spectral moment, which is the conditional spectral variance, and introduced the notion of the instantaneous bandwidth of a signal [6–9]. In particular, Cohen has argued that the second conditional spectral moments are

$$\langle \omega^2 \rangle_t = \left( \frac{A'(t)}{A(t)} \right)^2 + \varphi'^2(t) \quad (3)$$

$$\mu_\omega^2(t) = \sigma_\omega^2(t) = \langle \omega^2 \rangle_t - \langle \omega \rangle_t^2 = \left( \frac{A'(t)}{A(t)} \right)^2, \quad (4)$$

where the square root of the latter quantity is defined as the instantaneous bandwidth of the signal. (Poletti presents an alternative interesting viewpoint of instantaneous bandwidth, and has shown how to derive it in terms of a local Taylor series expansion of the signal [17]. In particular, a first-order expansion yields Cohen's definition.) Instantaneous bandwidth, like instantaneous frequency, is an important physical quantity that characterizes time-varying spectral properties of the signal, and has found application in a variety of areas, including acoustics, Doppler flow measurements, and seismology [1, 3, 15].

In this paper, we extend these ideas to higher instantaneous moments, such as the instantaneous skew and the instantaneous kurtosis, which are third and fourth order moments, respectively [11, 16]. Building on the foundations laid by Ville and Cohen, we derive expressions for all of the instantaneous moments of a signal, with particular attention to the skew and kurtosis.

## 2. BACKGROUND

The procedure we use is to apply operator methods to derive the instantaneous moments, analogous to the approach used by Cohen [6]. In the context of time-frequency analysis, Ville was the first to use the operator method, which was subsequently significantly extended by Cohen; in particular, Cohen has shown how

to obtain local expectation values (mean and variance) using operator methods [5, 6].

The fundamental idea is that the physical quantity of interest (in our case, frequency,  $\omega$ ) is represented by a Hermitian operator (for frequency the operator is  $\mathcal{W} = \frac{1}{j} \frac{d}{dt}$  in the time domain). A key aspect of the operator method is that one can obtain frequency averages directly from the signal, without having to first obtain the Fourier transform of the signal,  $S(\omega) = \frac{1}{\sqrt{2\pi}} \int s(t) e^{-j\omega t} dt$ . For example, the average value of  $\omega$  and of  $\omega^2$  (the first two global spectral moments) are given by [6, 7]<sup>1</sup>

$$\begin{aligned} \langle \omega \rangle &= \int \omega |S(\omega)|^2 d\omega = \int s^*(t) \mathcal{W} s(t) dt \\ &= \int \frac{\mathcal{W} s(t)}{s(t)} |s(t)|^2 dt \end{aligned} \quad (5)$$

$$\begin{aligned} \langle \omega^2 \rangle &= \int \omega^2 |S(\omega)|^2 d\omega = \int s^*(t) \mathcal{W}^2 s(t) dt \\ &= \int \left| \frac{\mathcal{W} s(t)}{s(t)} \right|^2 |s(t)|^2 dt, \end{aligned} \quad (6)$$

where the rules of operator algebra were used in arranging the latter term of eq. (6) to show the inherent positivity of the second moment (equivalently, one may substitute in the operator and use integration by parts to obtain a positive integrand [14]).

Since one is averaging over time in the expressions above to obtain the global average, Cohen and others have reasoned that the integrand represents the instantaneous value of the quantity [6, 7, 18]. In particular, for the signal  $s(t) = A(t)e^{j\varphi(t)}$ , we have

$$\frac{\mathcal{W} s(t)}{s(t)} = \varphi'(t) - j \frac{A'(t)}{A(t)}. \quad (7)$$

In the method of Cohen [6, 7], the average value of  $\omega$  at a given time is obtained from the real part (recall that the operator is Hermitian, and therefore the imaginary term integrates, per eq. (5), to zero),

$$\langle \omega \rangle_t = \left( \frac{\mathcal{W} s(t)}{s(t)} \right)_R = \varphi'(t), \quad (8)$$

which is the instantaneous frequency. Substituting this result into (5) gives the well-known result derived by Ville [18], namely that the time-average of the instantaneous frequency equals the global average frequency. Hence the interpretation of the (real part of the) integrand in (5) as the instantaneous (first) moment of the signal (times the magnitude-square of the signal).

Analogously, Cohen has derived the second moment [6–9]; following from eqs. (6) and (7), we have that the instantaneous second spectral moment is given by,

$$\langle \omega^2 \rangle_t = \left| \frac{\mathcal{W} s(t)}{s(t)} \right|^2 = \left( \frac{A'(t)}{A(t)} \right)^2 + \varphi'^2(t). \quad (9)$$

From this, Cohen has defined the instantaneous bandwidth of a signal, which is given by the square-root of the conditional variance,

$$\sigma_\omega^2(t) = \langle \omega^2 \rangle_t - \langle \omega \rangle_t^2 = \left( \frac{\mathcal{W} s(t)}{s(t)} \right)_I^2 = \left( \frac{A'(t)}{A(t)} \right)^2. \quad (10)$$

Note that Cohen's method gives an instantaneous variance (and bandwidth) that is positive, which is necessary for a proper interpretation. Also we re-write Cohen's instantaneous bandwidth expression equivalently as

$$\mu_\omega^2(t) = \sigma_\omega^2(t) = \left| \frac{(\mathcal{W} - \langle \omega \rangle_t) s(t)}{s(t)} \right|^2, \quad (11)$$

to highlight the fact that it is a central moment. This form will be convenient in the next section where we derive the instantaneous spectral kurtosis, which is a fourth order central moment, of the signal.

### 3. HIGHER INSTANTANEOUS SPECTRAL MOMENTS

We begin by presenting an identity that will be fundamental to our derivations, and which is a generalization of the central moment expression above for instantaneous bandwidth. Specifically, it can be shown that [12]

$$(\mathcal{W} - \varphi'(t))^n s(t) = (-j)^n A^{(n)}(t) e^{j\varphi(t)}, \quad (12)$$

where  $A^{(n)}(t)$  denotes  $\left(\frac{d}{dt}\right)^n A(t)$ .

The proof follows by induction. First, we show that the above relation holds for  $n=1$ ,

$$\begin{aligned} (\mathcal{W} - \varphi'(t)) s(t) &= \mathcal{W} s(t) - \varphi'(t) s(t) \\ &= -j \frac{d}{dt} (A(t) e^{j\varphi(t)}) - \varphi'(t) A(t) e^{j\varphi(t)} \\ &= -j A'(t) e^{j\varphi(t)}. \end{aligned} \quad (13)$$

Given that the identity holds for  $n=1$ , we proceed with the induction proof by assuming it holds for some  $n > 1$ , and then show that it holds for  $n+1$ , as follows:

$$\begin{aligned} &(\mathcal{W} - \varphi'(t))^{n+1} s(t) \\ &= (\mathcal{W} - \varphi'(t)) (\mathcal{W} - \varphi'(t))^n s(t) \\ &= (\mathcal{W} - \varphi'(t)) [(-j)^n A^{(n)}(t) e^{j\varphi(t)}] \end{aligned}$$

<sup>1</sup>Throughout the paper, the signal is normalized to unit-energy.

$$\begin{aligned}
&= (-j)^{n+1} \frac{d}{dt} \left( A^{(n)}(t) e^{j\varphi(t)} \right) \\
&\quad - (-j)^n \varphi'(t) A^{(n)}(t) e^{j\varphi(t)} \\
&= (-j)^{n+1} A^{(n+1)}(t) e^{j\varphi(t)}, \quad (14)
\end{aligned}$$

which completes the proof. We now derive expressions for the instantaneous moments of a signal.

### 3.1. Non-Central Moments

Like Cohen's instantaneous bandwidth above, higher-order even moments must be positive for a proper interpretation. By the operator procedure, the fourth order global moment may be written as

$$\begin{aligned}
\langle \omega^4 \rangle &= \int s^*(t) \mathcal{W}^4 s(t) dt = \int (\mathcal{W} s(t))^* \mathcal{W}^3 s(t) dt \\
&= \int (\mathcal{W}^2 s(t))^* \mathcal{W}^2 s(t) dt, \quad (15)
\end{aligned}$$

where we have used operator algebra to manipulate the integrands. We note that the three integrals are all equivalent. However, because the frequency operator does not commute with time, the three integrands are different, giving us (at least) three possible expressions for the fourth-order instantaneous moment, the averages of which all give the correct global moment, as is required. Only the integrand of the latter expression is positive, and thus we define the fourth-order instantaneous spectral moment as the integrand of the equation,

$$\langle \omega^4 \rangle = \int (\mathcal{W}^2 s(t))^* \mathcal{W}^2 s(t) dt = \int \left| \frac{\mathcal{W}^2 s(t)}{s(t)} \right|^2 |s(t)|^2 dt. \quad (16)$$

Specifically, we have that the instantaneous fourth order spectral moment is given by,

$$\begin{aligned}
\langle \omega^4 \rangle_t &= \left| \frac{\mathcal{W}^2 s(t)}{s(t)} \right|^2 \\
&= \left( \varphi'^2(t) - \frac{A''(t)}{A(t)} \right)^2 + \left( \varphi''(t) + 2 \frac{A'(t)}{A(t)} \varphi'(t) \right)^2. \quad (17)
\end{aligned}$$

This approach generalizes to higher-order even moments as [11, 16],

$$\langle \omega^{2m} \rangle_t = \left| \frac{\mathcal{W}^m s(t)}{s(t)} \right|^2. \quad (18)$$

Higher-order odd moments present an added challenge because, as with the even-order moments there are many different operator expressions that are possible, all of which integrate to give the correct global

moment, as is necessary; however, there is no constraint like positivity that we can use to specify a unique instantaneous moment. So for example, the possible third-order moments are given by the integrands of the following expression:

$$\langle \omega^3 \rangle = \int s^*(t) \mathcal{W}^3 s(t) dt = \int (\mathcal{W} s(t))^* \mathcal{W}^2 s(t) dt. \quad (19)$$

However, this ambiguity is resolved when we consider the central instantaneous moments, which we do next.

### 3.2. Central Moments

The instantaneous central moments are obtained by replacing the operator  $\mathcal{W}$  by  $\mathcal{W} - \langle \omega \rangle_t$  in the expressions for the non-central moments above [11, 16]. Doing so, it follows directly from eq. (18) that the even-order  $2m$ -th central instantaneous moment may be written as,

$$\mu_{\omega}^{2m}(t) = \left| \frac{(\mathcal{W} - \langle \omega \rangle_t)^m s(t)}{s(t)} \right|^2. \quad (20)$$

For the odd-order moments, we make use of the identity given in eq. (12), and the fact that the odd moments are obtained from the real part of the expression (analogous to the case for instantaneous frequency, which is a first order moment). It therefore follows immediately that the odd-order central instantaneous moments are zero, since for odd  $n$ , eq. (12) is purely imaginary. For example, the third-order instantaneous central moment is given by,

$$\mu_{\omega}^3(t) = \begin{cases} \left( \frac{(\mathcal{W} - \varphi'(t)) s(t)}{s(t)} \right)_R \\ \text{or} \\ \left( \frac{[(\mathcal{W} - \varphi'(t)) s(t)]^* (\mathcal{W} - \varphi'(t))^2 s(t)}{|s(t)|^2} \right)_R \end{cases} \quad (21)$$

$$= \begin{cases} \left( j \frac{A'''(t)}{A(t)} \right)_R \\ \text{or} \\ \left( -j \frac{A'(t) A''(t)}{A^2(t)} \right)_R, \end{cases} \quad (22)$$

which follows from the integrands of eq. (19) by substituting  $\mathcal{W} - \langle \omega \rangle_t$  for  $\mathcal{W}$  and writing them in the form of (21) times  $|s(t)|^2$ . Since we take the real part, all of these expressions evaluate to zero.

The fact that odd order central moments are identically zero by this method fixes the odd order non-central moments. In particular, we may use the binomial expansion,  $(a + b)^n = \sum_{k=0}^n \binom{n}{k} a^{n-k} b^k$ , to express

the central moments of the time-frequency distribution (eq. (2)) in terms of the non-central moments. For example, for  $n=3$ , we have that the third-order non-central instantaneous moment is given by,

$$\langle \omega^3 \rangle_t = \mu_\omega^3(t) + 3\langle \omega^2 \rangle_t \langle \omega \rangle_t - 2\langle \omega \rangle_t^3. \quad (23)$$

Since  $\mu_\omega^3(t) = 0$  and  $\langle \omega^2 \rangle_t$  and  $\langle \omega \rangle_t$  are unique, we obtain a unique third-order instantaneous non-central moment,

$$\langle \omega^3 \rangle_t = 3 \left( \frac{A'(t)}{A(t)} \right)^2 \varphi'(t) + \varphi'^3(t). \quad (24)$$

Similarly, the higher-order non-central odd moments can be uniquely expressed in terms of the central moment and lower-order moments.

From these results, we have immediately that the "instantaneous spectral skew," which we define analogously to its definition in ordinary densities, namely as a ratio of the third central moment to the second central moment, is zero,

$$\gamma_\omega(t) = \frac{\mu_\omega^3(t)}{(\mu_\omega^2(t))^{\frac{3}{2}}} = 0. \quad (25)$$

This result can be viewed as a generalization of the (global) spectral skew of the signal  $A(t)e^{j\omega_0 t}$ , which is zero since the spectrum is symmetric about the frequency  $\omega_0$ . Zero instantaneous spectral skew would occur if the instantaneous spectrum was symmetric about the instantaneous frequency  $\varphi'(t)$ .

The instantaneous kurtosis, which we again define analogously to its definition in ordinary probability, namely as a ratio of the fourth central moment to the second central moment, is given by

$$\begin{aligned} \kappa_\omega(t) &= \frac{\mu_\omega^4(t)}{(\mu_\omega^2(t))^2} \\ &= \left| \frac{(\mathcal{W} - \langle \omega \rangle_t)^2 s(t)}{s(t)} \right|^2 \bigg/ \left| \frac{(\mathcal{W} - \langle \omega \rangle_t) s(t)}{s(t)} \right|^4 \\ &= \frac{|s(t) (\mathcal{W} - \langle \omega \rangle_t)^2 s(t)|^2}{|(\mathcal{W} - \langle \omega \rangle_t) s(t)|^4} \\ &= \frac{(A(t)A''(t))^2}{A'^4(t)}, \end{aligned} \quad (26)$$

which follows from eqs. (11) and (20).

#### 4. EXAMPLES

##### 4.1. Identical Low-Order Moments for Different Signals

Just as different densities can have identical mean and variance, it is possible that two different signals can

have the same instantaneous frequency and bandwidth—but will have different instantaneous kurtosis. For example, consider the signals  $s_1(t) = A_1(t)e^{j\varphi(t)}$  and  $s_2(t) = A_2(t)e^{j\varphi(t)}$ , where  $A_2(t) = \frac{1}{A_1(t)}$  and  $\varphi(t)$  is an arbitrary phase function. The signals  $s_1(t)$  and  $s_2(t)$  obviously have identical instantaneous frequency. They also have identical instantaneous bandwidth since

$$\left( \frac{A'_2(t)}{A_2(t)} \right)^2 = \left( \frac{\frac{d}{dt} \frac{1}{A_1(t)}}{\frac{1}{A_1(t)}} \right)^2 = \left( \frac{-\frac{A'_1(t)}{A_1^2(t)}}{\frac{1}{A_1(t)}} \right)^2 = \left( \frac{A'_1(t)}{A_1(t)} \right)^2. \quad (27)$$

The instantaneous kurtosis (eq. (26)) is, however, different for each signal; in particular, the fourth central moment for  $s_2(t)$  is, by eqs. (20) and (12),

$$\begin{aligned} \mu_\omega^4(t) &= \left( \frac{A''_2(t)}{A_2(t)} \right)^2 = \left( A_1(t) \frac{d^2}{dt^2} \frac{1}{A_1(t)} \right)^2 \\ &= \left( 2 \left( \frac{A'_1(t)}{A_1(t)} \right)^2 - \frac{A''_1(t)}{A_1(t)} \right)^2 \neq \left( \frac{A''_1(t)}{A_1(t)} \right)^2. \end{aligned} \quad (28)$$

Thus, the time-varying spectral differences between these two signals are reflected in the higher instantaneous spectral moments.

##### 4.2. Positive Time-Frequency Density with Prescribed Conditional Moments

It is possible to construct TFDs that yield these moments. As an example, consider the sinusoidal FM signal with Gaussian amplitude,

$$s(t) = \left( \frac{16}{\pi} \right)^{\frac{1}{4}} e^{-8(t-0.5)^2 + j(30\pi t^2 + 28\pi t - 3 \sin(6\pi t))}. \quad (29)$$

We employ a moment constrained weighted least-squares (WLS) algorithm [13] to construct a positive time-frequency density (TFD) [10] which gives eqs. (8), (9), (24) and (17) as its first- through fourth-order conditional moments. The resulting TFD is shown in figure 1. Figure 2 shows the conditional moments of the TFD (solid) plotted against the proposed moments (dashed).

#### 5. CONCLUSION

We have given expressions for the instantaneous spectral moments of a signal, which are generalizations of the ideas of instantaneous frequency and instantaneous bandwidth. A simple, fundamental relationship between the central instantaneous moments and the amplitude of a signal was given, from which one can then obtain specific moments, such as the instantaneous skew and the instantaneous kurtosis. As with instantaneous frequency and instantaneous bandwidth, the instantaneous skew and kurtosis may be viewed as conditional

spectral moments of the time-varying spectral density of the signal. Having an expression for the higher conditional moments allows us to construct the density, and to differentiate between signals with the same instantaneous frequency and bandwidth. We speculate that, as in other areas where higher moments have been found to be useful, the same may hold true for the instantaneous kurtosis and higher moments introduced here for time-varying, or nonstationary, signals.

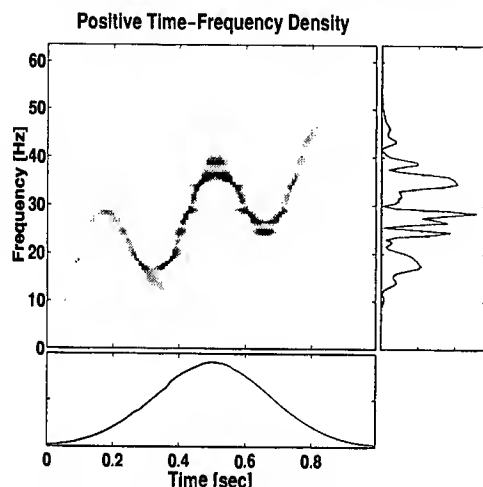


Figure 1: Positive TFD constrained to yield marginals and operator-derived moments for sinusoidal FM signal in (29). Side panel: frequency marginal. Bottom panel: time marginal.

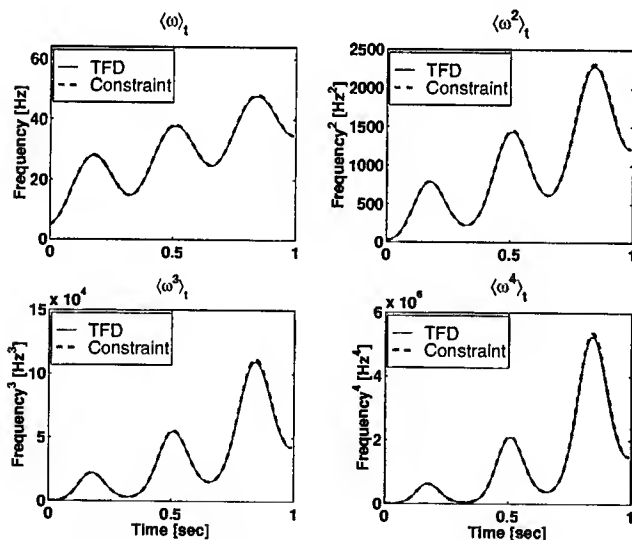


Figure 2: First- through fourth-order conditional moments of TFD in figure 1 (solid) constrained to yield the operator-derived moments (dashed).

## REFERENCES

- [1] A. Barnes, "Instantaneous Spectral Bandwidth and Dominant Frequency with Applications to Seismic Reflection Data," *Geophysics*, vol. 58, no. 3, pp. 418–428, March 1993.
- [2] B. Boashash, "Estimating and Interpreting the Instantaneous Frequency of a Signal—Part I: Fundamentals," *Proc. of the IEEE*, vol. 80, pp. 520–538, 1992.
- [3] J. Cardoso, M. Graça Ruano and P. Fish, "Nonstationarity Broadening Reduction in Pulsed Doppler Spectrum Measurements Using Time-Frequency Estimators," *IEEE Trans. Biomed. Eng.*, vol. 43, no. 12, pp. 1176–1186, 1996.
- [4] T. Claassen and W. Mecklenbräuker, "The Wigner Distribution—A Tool for Time-Frequency Signal Analysis—Part III: Relations with other Time-Frequency Signal Transformations," *Philips J. Research*, vol. 35, pp. 372–389, 1980.
- [5] L. Cohen, "Quantization Problem and Variational Principle in the Phase Space Formulation of Quantum Mechanics," *J. Math. Phys.*, vol. 17, pp. 1863–1866, 1976.
- [6] L. Cohen, "Instantaneous 'Anything'," *Proc. ICASSP '93*, vol. 4, pp. 105–108, 1993.
- [7] L. Cohen, *Time-Frequency Analysis*, Prentice-Hall, NJ 1995.
- [8] L. Cohen and C. Lee, "Instantaneous Frequency, Its Standard Deviation and Multicomponent Signals," *Proc. SPIE*, vol. 975, pp. 186–208, 1988.
- [9] L. Cohen and C. Lee, "Standard Deviation of Instantaneous Frequency," *Proc. ICASSP '89*, vol. 4, pp. 2238–2241, 1989.
- [10] L. Cohen and T. Posch, "Positive Time-frequency Distribution Functions," *IEEE Trans. Acous., Speech, and Sig. Proc.*, vol. 33, no. 1, pp. 31–38, 1985.
- [11] K. Davidson and P. Loughlin, "Instantaneous Spectral Moments," *J. Franklin Inst.*, (to appear).
- [12] K. Davidson, "Instantaneous Moments of a Signal," Ph.D. Thesis, University of Pittsburgh, 2000.
- [13] M. Emresoy and P. Loughlin, "Weighted Least-Squares Implementation of Cohen-Posch Time-Frequency Distributions with Specified Conditional and Joint Moment Constraints," *IEEE Trans. Sig. Proc.*, vol. 47, no. 3, pp. 893–900, 1999.
- [14] D. Gabor, "Theory of Communication," *IEE J. Comm. Engrng.*, vol. 93, pp. 429–441, 1946.
- [15] D. Hughes and L. Cohen, "Instantaneous Bandwidth and Local Duration in Acoustic Scattering," *Applied Signal Processing*, vol. 3, no. 2, pp. 68–77, 1996.
- [16] P. Loughlin and K. Davidson, "Instantaneous Kurtosis," *IEEE Sig. Proc. Ltr.*, vol. 7, no. 6, June 2000.
- [17] M. Poletti, "The Development of Instantaneous Bandwidth via Local Signal Expansion," *Sig. Proc.*, vol. 31, no. 3, pp. 273–281, 1993.
- [18] J. Ville, "Theorie et applications de la notion de signal analytique," *Cables et Transmissions*, vol. 2A, no. 1, pp. 61–74, 1948. [English Translation by I. Selin, "Theory and applications of the notion of complex signal," RAND Corp. Tech. Rep. T-92, Santa Monica, CA, 1958.]

# ADAPTIVE TIME-FREQUENCY REPRESENTATIONS FOR MULTIPLE STRUCTURES

*Antonia Papandreou-Suppappola*

Dept. of Electrical Engineering  
Arizona State University  
Tempe, AZ 85287-7206

E-mail: papandreou@asu.edu  
Web: www.eas.asu.edu/~apapand

*Seth B. Suppappola*

Pipeline Technologies, Inc.  
Scottsdale, AZ 85257-3773  
E-mail: seth@pipetech.com

## ABSTRACT

We propose an adaptive quadratic time-frequency representation (QTFR) based on a matching pursuit signal decomposition that uses a dictionary with elements matched to the instantaneous frequency of the analysis signal components. We form the QTFR as a weighted linear superposition of QTFRs chosen by the algorithm to provide a highly localized representation for each of the adaptively selected dictionary elements. This is advantageous as the resulting representations are parsimonious and reduce the effect of cross terms. Also, they exhibit maximum time-frequency localization for the difficult analysis case of signals with multiple components that have different time-frequency characteristics. Thus, the new technique can be used to analyze and classify multi-structure signal components as demonstrated by our synthetic and real data simulation examples.

## 1. INTRODUCTION

Nonstationary signals have been successfully analyzed using quadratic time-frequency representations (QTFRs) as they provide important information on the signals' time-varying characteristics [1–3]. Many QTFRs are ideally matched to one or two time-frequency (TF) structures based on the properties they satisfy. For example, Cohen's class QTFRs with signal-independent kernels [1,3] are matched to signals with linear TF characteristics as they preserve the signal's constant TF shifts. Hyperbolic or power QTFRs [4,5] are matched to signals with non-linear (dispersive) structures as they preserve dispersive time shifts. For successful TF analysis, it is important to match a QTFR with the TF structure of a signal. However, it is possible that a signal (for example, biological or sonar data) has multiple components with distinctively *different* TF structures. This complicates TF analysis due to the presence of cross terms or the effect of smoothing [3] that may impede interpretation.

Some QTFRs used for the analysis of signals with multiple TF structures include the spectrogram [3], reassigned QTFRs [6], and various adaptive QTFRs [7]. Although they work well in many applications, these QTFRs are not designed to yield the exact instantaneous frequency (IF) of a signal for classification. Also, they may not provide a well-localized representation without cross terms for analyzing

non-linear signal structures. Thus, it is very advantageous to design an adaptive QTFR to exactly match various signal components since many natural or synthetic signals may have different linear or non-linear TF structures. In this paper, we propose a *new* QTFR that adapts to different TF signal structures based on a matching pursuit algorithm.

## 2. BACKGROUND AND OBJECTIVES

The matching pursuit iterative algorithm of Mallat and Zhang decomposes a signal into a linear expansion of waveforms selected from a redundant and complete dictionary [8]. It uses successive approximations of the signal with orthogonal projections on dictionary elements. The dictionary consists of a basic Gaussian atom that is TF shifted and scaled. A QTFR (called the *modified Wigner distribution* in [8]) is obtained as a weighted superposition of the Wigner distribution (WD) [1–3] of each selected element. This QTFR is free of cross terms, and preserves signal energy, TF shifts, and scale changes on the analysis signal. It is also similar to a QTFR obtained in [9]. When a signal has multiple components with different TF structures, the QTFR uses many Gaussian elements to approximate the IF of each signal component. In order to analyze linear frequency-modulated (FM) chirps more efficiently with fewer waveforms, rotated Gaussian atoms were included in the dictionary in [10]. On the other hand, a wave-based dictionary consisting of wavefronts, resonances, and linear FM chirps was used to process scattering data in [11].

We propose to use a matching pursuit with dictionary elements that are matched to the constant, linear, or non-linear TF structure of a signal. These waveforms include complex sinusoids with linear or non-linear phase function such as logarithmic and power. Our aim is to analyze signals that have multiple IF structures such as the different characteristic signature whistles from a group of dolphins [12], and various biomedical signals measured simultaneously. The advantage of using a dictionary that is matched to the analysis data is that only a small number of elements will be used to decompose the signal, and the algorithm is expected to give fast and parsimonious results. At each iteration of the matching pursuit, we will *adaptively* choose the best dictionary element, identify its TF structure, and compute its corresponding QTFR. The resulting

proposed QTFR is formed as a weighted linear superposition of QTFRs that were each chosen to appropriately match a selected dictionary element. Such an algorithm not only designs a QTFR that is adapted to multiple signal structures, but can also be used in classification, detection, and identification applications for signals with specific linear and/or non-linear (dispersive) IF components.

### 3. ADAPTIVE ANALYSIS OF SIGNALS WITH MULTIPLE TF STRUCTURES

#### 3.1. Adaptive representation

Although the matching pursuit algorithm in [8] works well for many signals, we will show that it uses many Gaussian atoms to represent a signal component with non-linear TF characteristics. In addition, the modified WD is not very localized along the non-linear IF of the signal components.

In this paper, we modify the original matching pursuit in order to decrease the number of algorithm iterations, to improve the TF localization in the analysis of different multiple non-linear FMs, and to correctly classify the IF of each signal component. We follow the same basic concept of the matching pursuit algorithm in [8], but with some *major* differences. The first major difference is that we use more than one type of basic atom in our dictionary. Particularly, the dictionary consists of a large class of different basic atoms each of which has the form of a non-linear FM chirp

$$g(t; \xi, \lambda) = \sqrt{|\nu(t)|} e^{j2\pi \lambda \xi(\frac{t}{t_r})} \quad (1)$$

which is uniquely specified by its FM rate  $\lambda$  and its monotonic phase function  $\xi(b)$ . Note that  $\nu(t) = \frac{d}{dt} \xi(\frac{t}{t_r})$  is the IF of the signal in (1), and  $t_r > 0$  is a reference time. The dictionary may contain one type of FM chirp with fixed  $\xi(b)$  in (1) or a linear combination of them including sinusoids with  $\xi(b) = b$ , linear FM chirps with  $\xi(b) = \text{sgn}(b)|b|^2$ , hyperbolic FM chirps with  $\xi(b) = \ln b$ , power FM chirps with  $\xi(b) = \text{sgn}(b)|b|^\kappa$ , and exponential FM chirps with  $\xi(b) = e^b$ . The dictionary is formed by transforming the non-linear FM chirp<sup>1</sup> in (1) as

$$\begin{aligned} g(t; \xi, \lambda, \theta) &= \left( \mathcal{S}_\tau \mathcal{C}_a \mathcal{G}_c g(\xi, \lambda) \right)(t) \\ &= \sqrt{|a|} g(a(t - \tau); \xi, \lambda) e^{j2\pi c \xi(a(\frac{t-\tau}{t_r}))} \\ &= \sqrt{|a \nu(a(t - \tau))|} e^{j2\pi (c + \lambda) \xi(a(\frac{t-\tau}{t_r}))}, \end{aligned} \quad (2)$$

with the parameter vector  $\theta = [c, a, \tau] \in \Theta = \mathbb{R}^3$ . The unitary operators  $\mathcal{G}_c$ ,  $\mathcal{C}_a$ , and  $\mathcal{S}_\tau$  result in a constant FM rate shift  $c$ , scale change  $a$ , and constant time shift  $\tau$ , respectively, of the FM chirp. Specifically, the operators transform a signal  $x(t)$  as  $(\mathcal{G}_c x)(t) = x(t) e^{j2\pi c \xi(\frac{t}{t_r})}$ ,  $(\mathcal{C}_a x)(t) = \sqrt{|a|} x(at)$ , and  $(\mathcal{S}_\tau x)(t) = x(t - \tau)$ . Note that in (2), we use a transformation that results in a constant shift (from  $\lambda$  to  $c + \lambda$ ) of the FM rate of the non-linear FM chirp instead of a constant frequency shift as in [8]. This is because we are considering signals that may be wide-band as well as dispersive, thus a shift of the IF is a better matched transformation to cover the entire TF plane [4].

<sup>1</sup>Without loss of generality, the atom in (1) may use  $\lambda = 1$ .

With appropriate normalization, we restrict the energy of  $g(t; \xi, \lambda, \theta)$  to be unity for every  $\theta$  in order to ensure energy preservation when  $\xi(b)$  is fixed [8, 13]. The iterative procedure of the matching pursuit first projects the analysis signal  $x(t) = (R_0 x)(t)$  onto each element of the dictionary, and selects<sup>2</sup>  $g(t; \xi_0, \lambda, \theta_0)$  based on the condition<sup>3</sup>  $|(x, g(\xi_0, \lambda, \theta_0))| \geq |(x, g(\xi, \lambda, \theta))|, \forall \theta \in \Theta$  and for all possible phase functions  $\xi(b)$  of the elements used to form the dictionary. This ensures that the element with the highest energy will be chosen first. This results in the signal decomposition

$$x(t) = \beta_0 g(t; \xi_0, \lambda, \theta_0) + (R_1 x)(t) \quad (3)$$

with the expansion coefficient  $\beta_0 = (x, g(\xi_0, \lambda, \theta_0))$ . The function  $\xi_0(b)$  corresponds to the phase function of this first, highest energy signal component of the analysis signal. For example, if the first dictionary element chosen is a hyperbolic FM chirp, then  $\xi_0(b) = \ln b$ .

The second major difference of our algorithm from the one in [8–10] is that we do not compute the WD of each selected element to form the modified WD. Instead, we adaptively use the information that the first selected waveform has phase function  $\xi_0(b)$  in order to compute its *generalized warped Wigner distribution* (GWD) [5, 14]. The GWD is a warped version of the WD, with the warping [5, 14, 15] based on a monotonic and (possibly) non-linear parameter function  $\zeta(b)$ . In particular,

$$\text{GWD}_x(t, f; \zeta) = \text{WD}_y \left( t_r \zeta \left( \frac{t}{t_r} \right), \frac{f}{t_r \mu(t)} \right) \quad (4)$$

where  $\mu(t) = \frac{d}{dt} \zeta \left( \frac{t}{t_r} \right)$ , and the warped signal is [5]

$$y(t) = (\mathcal{W}_\zeta x)(t) = t_r |\mu(t_r \zeta^{-1}(\frac{t}{t_r}))|^{-1/2} x(t_r \zeta^{-1}(\frac{t}{t_r})).$$

Note that a specific GWD is obtained simply by fixing its parameter function  $\zeta(b)$ . By matching  $\zeta(b)$  in (4) to be equal to the phase function  $\xi_0(b)$  in (3) (i.e. if  $\zeta(b) = \xi_0(b)$ ), our new *adaptive representation for multiple structures* (ARMUS) QTFR, at this first iteration, is simply

$$T_x^0(t, f) = |\beta_0|^2 \text{GWD}_{g(\xi_0, \lambda, \theta_0)}(t, f; \xi_0).$$

At the second iteration, the residual function  $(R_1 x)(t)$  is obtained by solving (3), and it is decomposed in a similar manner to the signal  $x(t)$ . At the  $(n + 1)$ th iteration, the criterion

$$|(R_n, g(\xi_n, \theta_n))| \geq |(R_n, g(\xi, \theta))|, \quad \forall \theta \in \Theta \quad (5)$$

is used to decompose the  $n$ th residual function  $(R_n x)(t)$  as  $(R_n x)(t) = \beta_n g(t; \xi_n, \theta_n) + (R_{n+1} x)(t)$  where

$$\beta_n = (R_n x, g(\xi_n, \theta_n)) \quad (6)$$

is the expansion coefficient. The GWD of  $(R_n x)(t)$  is also obtained adaptively to match the TF structure of the  $n$ th residual function by letting  $\zeta(b) = \xi_n(b)$  in (4).

<sup>2</sup>Note that a subscript  $n$  in the parameters  $R_n$ ,  $\xi_n(b)$ ,  $\theta_n$ ,  $\tau_n$ , and  $c_n$ , and a superscript  $n$  in a QTFR  $T^n(t, f)$  indicate the algorithm parameters at the  $(n + 1)$ th iteration.

<sup>3</sup>The inner product is defined as  $\langle x, g \rangle = \int_{-\infty}^{\infty} x(t) g^*(t) dt$ .



After a total of  $N$  iterations, the matching pursuit algorithm results in the signal decomposition

$$x(t) = \sum_{n=0}^{N-1} \beta_n g(t; \xi_n, \lambda, \underline{\theta}_n) + (R_N x)(t). \quad (7)$$

Note that since our dictionary is complete,<sup>4</sup> any signal  $x(t)$  can be represented as in (7) with  $N = \infty$ , in which case  $(R_N x)(t) = 0$  [8]. In actuality, when the signal components match the TF structure of the dictionary elements, the algorithm converges quickly. As stopping criteria, we use a maximum number of iterations, and an acceptable small residue energy compared to the data energy [8].

The resulting ARMUS of the signal at the  $N$ th iteration is the weighted sum of the appropriate GWDs of the dictionary elements selected at each iteration. Specifically,

$$\begin{aligned} \text{ARMUS}_x(t, f) &= T_x^{N-1}(t, f) \\ &= \sum_{n=0}^{N-1} |\beta_n|^2 \text{GWD}_{g(\xi_n, \underline{\theta}_n)}(t, f; \xi_n), \end{aligned} \quad (8)$$

with the weights  $|\beta_n|^2$  defined in (6). Note that the same GWD (with fixed parameter function  $\xi_n(b)$ ) in (8) may be used if multiple chirps have the same phase function but different FM rate.

### 3.2. Decomposition and QTFR properties

An important property of the matching pursuit in (7) is its covariance to certain signal changes. Consider the decomposed signal  $x(t) = \sum_{n=0}^{\infty} \beta_n g(t; \xi, \lambda, \underline{\theta}_n)$  in (7) with  $N = \infty$ , and with similar TF structure dictionary elements (i.e. let  $\xi_n(b) = \xi(b)$ ,  $\forall n$ ). If the FM rate of a non-linear chirp  $x(t)$  is shifted by a constant amount to form  $y(t) = (\mathcal{G}_u x)(t) = x(t) e^{j2\pi u \xi(\frac{t}{\tau_n})}$ , then its matching pursuit is simply given as  $y(t) = \sum_{n=0}^{\infty} \beta_n g(t; \xi, \lambda, \underline{\theta}_n)$ . Note that the expansion coefficients  $\beta_n$  are not affected by this signal change. The parameter vector changes to  $\underline{\theta}_n = [c_n + u, a_n, \tau_n]$  indicating that the time shifts  $\tau_n$  and the scale changes  $a_n$  remain the same, whereas the dictionary atoms undergo a constant shift in their FM rate from  $(\lambda + c_n)$  to  $(\lambda + c_n + u)$ . Note that if  $\xi(b)$  is a power or a logarithmic function, then we can show that the corresponding matching pursuit is also covariant to scale changes [13].

The ARMUS QTFR in (8) also satisfies various properties that are desirable in many applications. By simply combining the GWDs of each selected dictionary element, no cross terms are introduced in the QTFR. Also, it preserves the underlying TF structure of each analysis signal component, and it provides a highly localized representation of each component as it does not apply any smoothing. Specifically, the GWD with parameter  $\xi(b)$  of a non-linear FM chirp with IF  $\xi(b)$  results in the highly localized representation  $\text{GWD}_{g(\xi, \lambda)}(t, f; \xi) = |\nu(t)| \delta(f - \lambda \nu(t))$  [5]. If a particular application uses signal components with only one type of TF structure, then we should form our dictionary using the corresponding non-linear FM chirp with matched IF. In such cases, the ARMUS satisfies other desirable signal properties such as the preservation of signal energy, and

changes in the analysis signal's FM rate [13]. If the dictionary elements are either hyperbolic or power FM chirps, then the QTFR also preserves scale changes. In [9], it was shown that if some cross terms are allowed in a version of the modified WD, then additional signal properties, such as the marginals, can be satisfied. This depends on a distance measure criterion that controls the amount of cross terms included in the QTFR formulation. We are currently investigating the corresponding distance measure for each different dispersive QTFR function  $\xi(b)$ .

### 3.3. Implementation issues

As we vary many parameters in our algorithm in order to select the appropriate dictionary elements for the matching pursuit, the computation is intensive. However, if we preprocess our data, we can form a dictionary with elements which approximately span the data in TF structure. Thus, the algorithm iterates more rapidly. Additional speedup is possible if we compute the matched GWD of each dictionary element ahead of time. Since the last operation on the basic atoms in (2) is time shifting, we perform the inner products in the matching pursuit criterion in (5) as a cross-correlation instead of introducing another layer of dictionary elements over all possible time shifts. Thus, the inner products in (5) are computed as correlations between the residual functions and the dictionary elements that have been generalized frequency-shifted over all FM rates  $c$  and scaled over all  $a$ . This increases the computational speed since correlations can be implemented using the fast Fourier transform (FFT). Also, the memory consumption by the dictionary is significantly reduced since additional dictionary elements are not needed for every time shift. Moreover, since the dictionary elements do not change, and the residual data are constant during a given matching pursuit iteration, additional speedup could be achieved by pre-computing and storing the FFTs of these sequences.

If the signal components are well-separated in time, we can use the algorithm to simply find the time support and phase function of each selected element, and then use the information to analyze the actual data (instead of the selected waveforms) with its matched GWD. This will greatly reduce computation as only a few GWDs need to be obtained. If classification is needed without analysis, the algorithm can provide the IF of each signal component without computing its QTFR, simply by extracting that information from the matched dictionary elements.

### 3.4. Simulation examples

**Synthetic data:** We demonstrate the performance of our new QTFR by first analyzing a synthetic signal with seven components: two windowed hyperbolic FM chirps and five windowed linear FM chirps with different chirp rates, scalings, and time shifts. Their "ideal" TF structure obtained by adding the IF of each component is shown in Fig. 1(a). The WD in Fig. 1(b) suffers from cross terms and makes it difficult to identify the true TF structure of each component. On the other hand, the spectrogram [3] in Fig. 1(c) suffers from loss of resolution due to smoothing that prohibits signal classification and the identification of the exact number of signal terms.

<sup>4</sup>The proof can be found in [13].

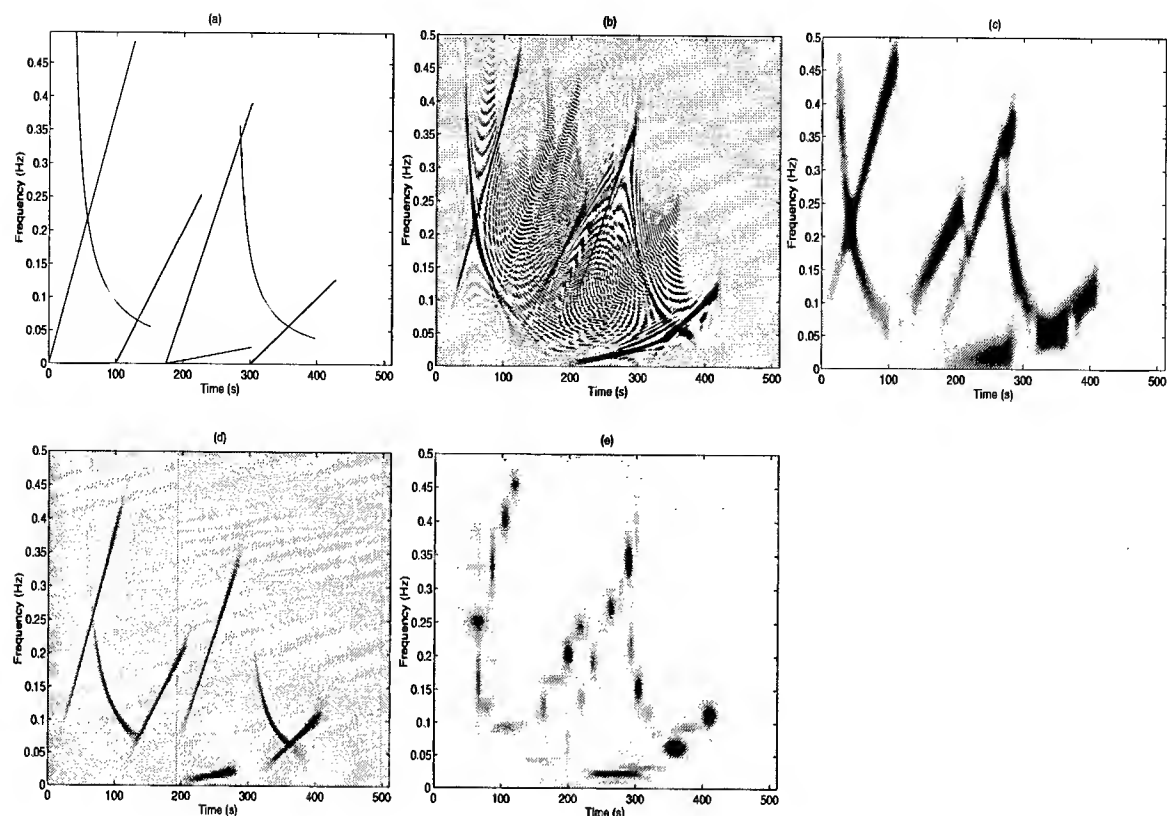


Figure 1: (a) A linear combination of the ideal IF of each component of a signal consisting of two windowed hyperbolic FM chirps and five windowed linear FM chirps. The signal is analyzed using (b) the Wigner distribution, (c) the spectrogram, (d) the new ARMUS QTFR, and (e) the modified WD in [8].

We apply our method by decomposing the signal using a dictionary of linear and hyperbolic FM chirps with about 89,000 combinations of FM rate changes, scale changes, and time shifts. The decomposition approximates the data very well after only *seven* iterations (the same number as the signal terms) as demonstrated by overlaying the signal with its expansion. The ARMUS QTFR in Fig. 1(d) provides a highly localized representation for all seven components without outer cross terms or loss of resolution. This is because it adaptively computes the Altes Q-distribution [4] for selected elements with hyperbolic TF characteristics, and the WD for selected elements with linear TF characteristics. Note that the mild spreading of the signal components is due to the fact that the data was windowed for processing. For further comparison, we applied the matching pursuit from [8] with Gaussian dictionary elements to decompose the signal, and then we analyzed it using the modified WD as shown in Fig. 1(e). Note that although the QTFR does not yield any cross terms, it does not provide a localized representation that can easily identify the TF structure of the components. Also, the algorithm does not converge with as many as fifty iterations, and it does not provide a closed form estimate of the IF of the signal components for classification.

**Real data:** We use our matching pursuit algorithm to obtain a closed form estimate of the IF of real data for clas-

sification. The analysis data consists of whistles<sup>5</sup> from a long-finned pilot whale. In Fig. 2(a), the spectrogram of the data shows three whistles with dispersive TF characteristics as high frequencies are time-delayed by a shorter amount than low ones. Although the spectrogram provides visual information, it cannot find the exact IF of the signal components. Our matching pursuit decomposition of the data is highly localized along hyperbolic TF curves as we formed our algorithm using hyperbolic FM chirps. This is shown by plotting the sum of the IFs of the selected waveforms in Fig. 2(b). Fig. 2(c) shows an overlay of the plots in Figs. 2(a) and 2(b) for a fair comparison. Note that based on the spectrogram analysis, we set the iteration limit to three. However, the algorithm did not extract the third component since (i) it is low in amplitude, and (ii) the higher frequency component is not exactly hyperbolic, so the algorithm keeps trying to remove that component first. On the other hand, the matching pursuit provided us with a *closed form estimate* of the true IF of the two louder whistles. For better classification, we plan to increase the number of iterations as well as include in our dictionary both hyperbolic and power FM chirps. We expect to obtain better matched results since the IF of the higher frequency whistle appears to be a power function.

<sup>5</sup>The data was obtained from the database of W. Watkins [16] at the Woods Hole Oceanographic Institute.

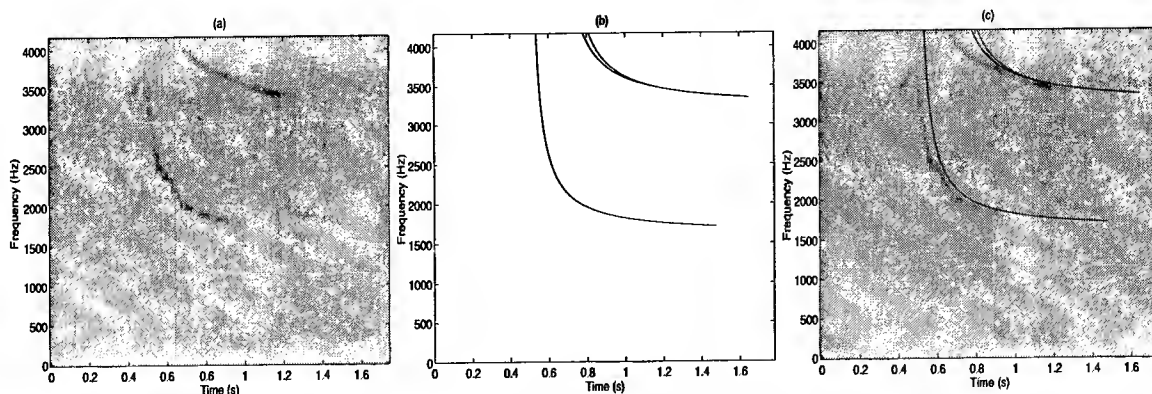


Figure 2: Analysis of real data whistles from a long-finned pilot whale: (a) spectrogram, (b) sum of IFs of selected waveforms from a matching pursuit with a hyperbolic FM chirp dictionary after three iterations, (c) an overlay of the first two plots.

#### 4. CONCLUSION

We have developed an adaptive QTFR based on the matching pursuit algorithm in order to analyze time-varying signals that have multiple components with different (possibly non-linear) frequency modulation. We build our dictionary based on pre-processing the analysis data to obtain some general information on the TF structure of the signal components. At each iteration, we adaptively match the selected dictionary element with the matched QTFR that provides its most localized representation. The resulting QTFR of the analysis signal is a linear superposition of the individual QTFRs of the elements, weighted by the magnitude squared of the expansion coefficients. We have demonstrated with simulated examples that this new QTFR handles well the difficult problem of analyzing signals with multiple IF structures in TF signal processing without introducing cross terms, or altering the underlying structure of each signal component, or suffering from a loss of resolution due to smoothing.

**Acknowledgement:** We would like to thank Dr. G. Faye Boudreaux-Bartels for valuable advice and initial collaboration on this topic.

#### REFERENCES

- [1] L. Cohen, *Time-Frequency Analysis*, Englewood Cliffs, NJ: Prentice Hall, 1995.
- [2] P. Flandrin, *Time-Frequency/Time-Scale Analysis*, San Diego, CA: Academic Press, 1999.
- [3] F. Hlawatsch and G. F. Boudreaux-Bartels, "Linear and quadratic time-frequency signal representations," *IEEE Signal Proc. Mag.*, vol. 9, pp. 21-67, April 1992.
- [4] A. Papandreou, F. Hlawatsch, and G. F. Boudreaux-Bartels, "The hyperbolic class of quadratic time-frequency representations, Part I," *IEEE Trans. on Signal Processing*, vol. 41, pp. 3425-3444, Dec. 1993.
- [5] A. Papandreou-Suppappola, "Generalized time-shift covariant quadratic time-frequency representations with arbitrary group delays," *29th Asilomar Conference on Signals, Systems, and Computers*, (Pacific Grove, CA), pp. 553-557, October 1995.
- [6] F. Auger and P. Flandrin, "Improving the readability of time-frequency and time-scale representations by the reassignment method," *IEEE Transactions on Signal Processing*, vol. 43, pp. 1068-1089, May 1995.
- [7] R. G. Baraniuk and D. L. Jones, "A signal-dependent time-frequency representation: Optimal kernel design," *IEEE Transactions on Signal Processing*, vol. 41, pp. 1589-1602, April, 1993.
- [8] S. G. Mallat and Z. Zhang, "Matching pursuits with time-frequency dictionaries," *IEEE Trans. on Signal Processing*, vol. 41, pp. 3397-3415, December 1993.
- [9] S. Qian and D. Chen, "Decomposition of the Wigner distribution and time-frequency distribution series," *IEEE Transactions on Signal Processing*, vol. 42, pp. 2836-2842, October 1994.
- [10] A. Bultan, "A four-parameter atomic decomposition of chirplets," *IEEE Trans. on Signal Processing*, vol. 47, pp. 731-745, March 1999.
- [11] M. R. McClure and L. Carin, "Matching pursuits with a wave-based dictionary," *IEEE Transactions on Signal Processing*, vol. 45, pp. 2912-2927, December 1997.
- [12] W. W. L. Au, *The Sonar of Dolphins*, Springer-Verlag, 1993.
- [13] A. Papandreou-Suppappola, Telecommunications Research Center Report TRC4-00-1, Dept. of Electrical Engineering, Arizona State University, April 2000.
- [14] A. Papandreou-Suppappola, R. L. Murray, B. Iem and G. F. Boudreaux-Bartels, "Generalized time-shift covariant QTFRs," submitted to *IEEE Trans. on Signal Processing*, Nov. 1998.
- [15] R. G. Baraniuk and D. L. Jones, "Unitary equivalence: A new twist on signal processing," *IEEE Trans. on Signal Processing*, vol. 43, pp. 2269-2282, October 1995.
- [16] W. A. Watkins, K. Fristrup, M. A. Daher, and T. Howald, "SOUND database of marine animal vocalizations structure and operations," Techn. Rept. WHOI-92-31, Woods Hole Oceanographic Inst., August 1992.

# A RESOLUTION PERFORMANCE MEASURE FOR QUADRATIC TIME-FREQUENCY DISTRIBUTIONS

Boualem Boashash and Victor Sucic

Signal Processing Research Centre  
Queensland University of Technology  
GPO Box 2434, Brisbane, Qld 4001, Australia  
e-mail: b.boashash@qut.edu.au, v.sucic@qut.edu.au

## ABSTRACT

This paper presents two novel results which are significant for the application of time-frequency signal analysis techniques to real life signals. First, we introduce a measure for comparing the resolution performance of TFDs in separating closely spaced components in the time-frequency domain. The measure takes into account key attributes of TFDs such as main-lobes, side-lobes, and cross-terms. The introduction of this measure is an improvement of current techniques which rely on visual inspection of plots.

The second result consists in proposing a methodology for designing high resolution quadratic TFDs for the time-frequency analysis of multicomponent signals when components are close to each other. A recently introduced TFD, the B-distribution, and its modified version are defined using this methodology.

Finally, the performance comparison of quadratic TFDs using the proposed resolution measure shows that the B-distribution outperforms existing quadratic TFDs in resolving closely spaced components in the time-frequency domain.

## 1. INTRODUCTION

This paper describes what we believe is the first attempt at providing an objective quantitative measure criterion for comparing the performance of quadratic time-frequency distributions (TFDs), in terms of resolution (separation of closely spaced components), when applied to the analysis of multicomponent signals.

Let us consider a multicomponent signal given by:

$$s(t) = s_1(t) + s_2(t) \quad (1)$$

where  $s_1(t)$  and  $s_2(t)$  are two parallel linear frequency modulated (LFM) signals of length  $N = 128$  and sampling frequency  $f_s = 1\text{Hz}$ . The frequency of the first component  $s_1(t)$  goes from  $0.15\text{Hz}$  to  $0.25\text{Hz}$ , while the frequency of the second component  $s_2(t)$  varies from  $0.2\text{Hz}$  to  $0.3\text{Hz}$ .

The multicomponent signal  $s(t)$  is represented in the time-frequency domain using the Wigner-Ville distribution (WVD), the spectrogram, the Choi-Williams distribution (CWD) [1], the Born-Jordan distribution [2], Zhao-Atlas-Marks (ZAM) distribution [3], and the recently introduced B-distribution [4, 5] (see Figure 1).

The desire to objectively compare the plots in Figure 1 motivated the need to define a quantitative performance measure for TFDs. The characteristics of TFDs that influence their resolution, such as energy concentration, mainlobes separation, sidelobes and cross-terms minimisation, are combined to define a quantitative measure criterion.

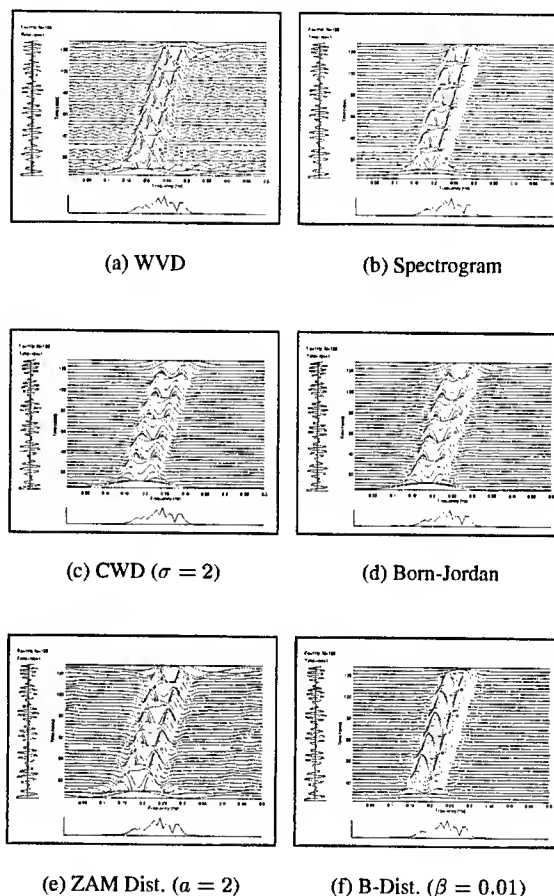


Figure 1: TFDs of two LFM signals with frequency  $f_1 = 0.15 - 0.25\text{Hz}$  and  $f_2 = 0.2 - 0.3\text{Hz}$ . All plots use a rectangular window, apart from the spectrogram which uses the Hanning window

This paper presents a comparison of the resolution performance of the above mentioned TFDs, using the newly proposed measure criterion. In this context, we show that the B-distribution outperforms the other quadratic TFDs for signals with components closely-spaced in the time-frequency plane.

## 2. PERFORMANCE CRITERIA OF TIME-FREQUENCY DISTRIBUTIONS

### 2.1. Monocomponent Signal

The performance of a TFD in the case of *monocomponent* FM signals is commonly defined in terms of the energy concentration the TFD achieves about the signal instantaneous frequency (IF) [6]. For the slice of TFD taken at the time instant  $t_0$ , illustrated in Figure 2, we may express the performance measure as:

$$p = \frac{|A_S|}{|A_M|} \frac{V}{f} \quad (2)$$

where  $A_M$  is the amplitude of the mainlobe of the TFD,  $A_S$  is the amplitude of the sidelobes,  $V$  is the 1.5 dB bandwidth<sup>1</sup> of the mainlobe and  $f$  represents the IF of the signal, all taken at time  $t_0$ . The rationale for introducing (2) is that one wants to minimise sidelobe amplitude  $A_S$  and mainlobe bandwidth  $V$  relative to central frequency  $f$ , but maximise mainlobe amplitude  $A_M$ .

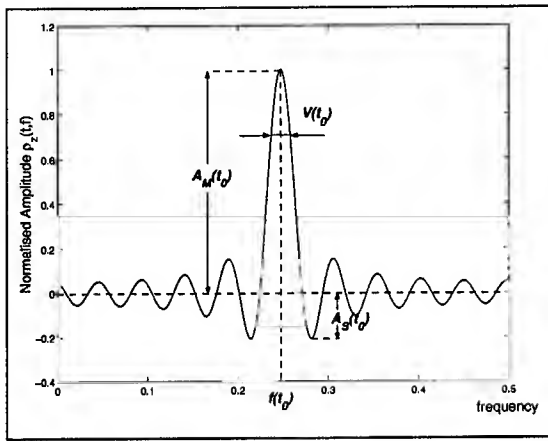


Figure 2: Slice of a TFD of a monocomponent signal taken at the time instant  $t = t_0$

### 2.2. Multicomponent Signal

The performance of time-frequency distributions of a *multicomponent* FM signal, can be *quantitatively* measured in terms of:

- the energy concentration of the distribution about the respective instantaneous frequency of each component, as expressed by equation (2), and
- the resolution as measured by the separation of the mainlobes of the components in the time-frequency plane, and the effect of cross-terms.

#### 2.2.1. Energy Concentration

By extending the concept in Section 2.1, a TFD is said to have the best energy concentration for a given multicomponent FM signal if for each of the signal components:

- its 1.5 dB mainlobe bandwidth relative to  $f$  is the smallest compared to that of other distributions, and if
- it yields the smallest sidelobe magnitude to mainlobe magnitude ratio compared to those of other distributions.

#### 2.2.2. Resolution

The frequency resolution in a power spectral estimate of a signal composed of two single tones,  $f_1$  and  $f_2$ , is defined as the minimum difference  $f_2 - f_1$  for which the following inequality holds:

$$f_1 + \frac{V_1}{2} < f_2 - \frac{V_2}{2} \quad (3)$$

where  $V_1$  and  $V_2$  are the 1.5 dB mainlobe bandwidth of the first and the second sinusoid, respectively, as illustrated in Figure 3.

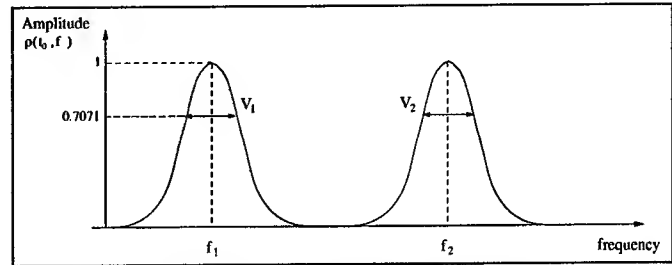


Figure 3: Resolution of a two-component signal

For a time-frequency distribution  $p_z(t, f)$  of a two-component signal, the above definition of resolution would be valid for every slice of cross-terms free TFDs, such as the spectrogram, taken at time  $t = t_0$ . However, for TFDs with cross-terms, we need to account for the effect of cross-terms on resolution, as illustrated by Figure 4 and explained in the next section.

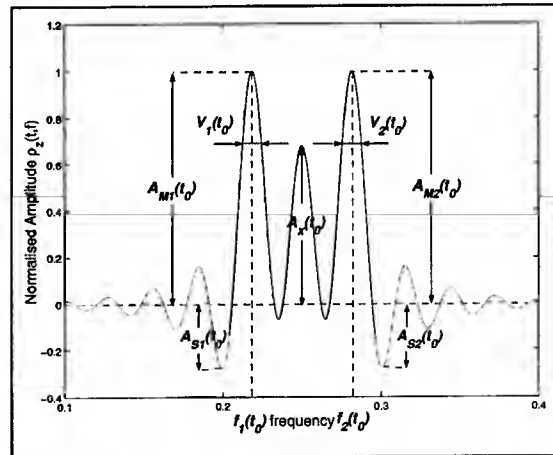


Figure 4: Slice of a TFD of a two-component signal taken at time  $t = t_0$

In Figure 4,  $V_1(t_0)$ ,  $f_1(t_0)$ ,  $A_{S1}(t_0)$  and  $A_{M1}(t_0)$  represent respectively the 1.5 dB mainlobe bandwidth, the instantaneous frequency, the sidelobe amplitude and the mainlobe amplitude of the

<sup>1</sup>We measure the bandwidth of the mainlobe of a component at the rms value of the component normalised amplitude. See also footnote 5.

first component at time  $t = t_0$ . Similarly,  $V_2(t_0)$ ,  $f_2(t_0)$ ,  $A_{S_2}(t_0)$  and  $A_{M_2}(t_0)$  represent the 1.5 dB mainlobe bandwidth, the instantaneous frequency, the sidelobe amplitude and the mainlobe amplitude of the second component at the same time  $t_0$ .  $A_X(t_0)$  defines the cross-terms amplitude.

### 2.2.3. Resolution Performance Measure of TFDs

Equation (3) and Figure 4 suggest that the resolution performance of a time-frequency distribution of a two-component signal is given by the minimum value of the difference  $R = f_2 - f_1$  for which we still have a positive separation  $D$  between the components' mainlobes about their respective IFs,  $f_2$  and  $f_1$ . For TFDs,  $D$  should ideally be as close as possible to the true difference between the actual frequencies. It is expressed as:

$$D = \frac{(f_2 - \frac{V_2}{2}) - (f_1 + \frac{V_1}{2})}{f_2 - f_1} = 1 - \frac{V_1 + V_2}{2R} \quad (4)$$

The resolution also depends on the following set of variables, all of which should be as small as possible:

- the 1.5 dB normalised *mainlobe bandwidth* of the signal component  $V_k/f_k$ ,  $k = 1, 2$ , which is already included in  $D$  (equation (4)),
- the ratio of the *sidelobe magnitude*  $|A_{S_k}|$  to the mainlobe magnitude  $|A_{M_k}|$ ,  $k = 1, 2$  of the components, and
- the ratio of the *cross-term magnitude*  $|A_X|$  to the mainlobe magnitude of the signal auto-terms  $|A_{M_k}|$ ,  $k = 1, 2$ .

It follows that the best TFD for *multicomponent signals analysis* is the one that **minimises** the positive quantities a), b) and c), and **maximises**<sup>2</sup> the separation  $D$ , concurrently.

Hence, an indicator  $P$  of the resolution performance of a given TFD can be defined as [7]:

$$P = \frac{|A_S||A_X|}{|A_M|^2 D} \geq 0 \quad (5)$$

where  $A_M$ ,  $A_S$  and  $A_X$  are respectively the average amplitudes of the mainlobes, sidelobes and cross-terms of any two consecutive components of the multicomponent signal, with  $D$  being their relative separation.

If  $P < 0$ , then there is no separation of the components, while if  $P \geq 0$ ,  $P$  provides a measure of the resolution performance, which takes into account separation  $D$  and the effect of cross-terms (best performance is achieved by minimising  $P$ ).

## 3. TIME-FREQUENCY SIGNAL ANALYSIS OF CLOSELY SPACED COMPONENTS USING THE B-DISTRIBUTION

### 3.1. Defining TFDs via Ambiguity Filtering

Different time-frequency distributions of the analytic signal  $z(t)$ , associated with the real signal  $s(t)$ , can be obtained by selecting different kernel functions  $g(\nu, \tau)$  in the general expression of the quadratic class<sup>3</sup> [8]:

$$\rho_z(t, f) = \iiint g(\nu, \tau) e^{j2\pi\nu(t-u)} z(u + \frac{\tau}{2}) z^*(u - \frac{\tau}{2}) e^{-j2\pi f\tau} d\nu d\tau \quad (6)$$

<sup>2</sup>The maximum value is  $D = 1$  which is obtained when  $V_1 = V_2 = 0$ .

<sup>3</sup>All three integrals have limits from  $-\infty$  to  $+\infty$ . Note: this formula differs from Cohen's formula by a minus sign in the first exponential.

For  $g(\nu, \tau) = 1$ , we obtain the Wigner-Ville distribution (WVD) of the signal [2, 9]:

$$WVD_z(t, f) = \int_{-\infty}^{\infty} z(t + \frac{\tau}{2}) z^*(t - \frac{\tau}{2}) e^{-j2\pi f\tau} d\tau \quad (7)$$

A key to understanding time-frequency relationships is through understanding of the ambiguity domain. The symmetrical ambiguity function (AF) is defined as:

$$AF_z(\nu, \tau) = \int_{-\infty}^{\infty} z(t + \frac{\tau}{2}) z^*(t - \frac{\tau}{2}) e^{-j2\pi\nu t} dt \quad (8)$$

From equations (7) and (8) we can see that the WVD and the AF are related by a two-dimensional Fourier transform [2]:

$$WVD_z(t, f) \overset{t}{\rightleftharpoons}_{\tau} AF_z(\nu, \tau)$$

$$WVD_z(t, f) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} AF_z(\nu, \tau) e^{-j2\pi(f\tau - \nu t)} d\nu d\tau \quad (9)$$

It was shown that a signal mapped by the AF into the Doppler-lag domain always traverses the origin of that plane, while the cross-terms, having oscillating amplitude in the time-frequency domain, are located away from the origin in the Doppler-lag plane, the distance being directly proportional to the time and frequency distance of the signal components [1].

This property of the AF has inspired researchers to look for two-dimensional kernel filters  $g(\nu, \tau)$  that enhance the generalised ambiguity function,  $g(\nu, \tau)AF_z(\nu, \tau)$ , around its origin and suppress it elsewhere.

Using equations (6) and (9), the following expression can also be derived [2]:

$$\rho_z(t, f) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(\nu, \tau) AF_z(\nu, \tau) e^{-j2\pi(f\tau - \nu t)} d\nu d\tau \quad (10)$$

Thus, quadratic TFDs may be found by filtering the symmetrical ambiguity function with  $g(\nu, \tau)$  and then carrying out the two-dimensional Fourier transform. For example, for the Wigner-Ville distribution with the ambiguity domain kernel filter equal to unity, no filtering is applied to the AF, resulting in the complete preservation of the cross-terms. This in return makes the interpretation of the WVD of multicomponent signals highly difficult. The spectrogram, on the other hand, leads to a quasi-total elimination of the cross-terms to the detriment of resolution.

### 3.2. New Constraints for TFD Design

It was reported in [2] that for a time-frequency analysis, a TFD is expected to be real, to satisfy the marginals and to have the instantaneous frequency as its first moment with respect to frequency. These strict constraints on the kernel design in the ambiguity domain [9] led to the terminology of Cohen's class.

However, it is known that the spectrogram does not exhibit cross-terms, and does not satisfy the marginals. Yet the spectrogram is a very popular tool in practical applications, suggesting that the time and the frequency marginal constraints may not be really strictly needed in practice. What may be more important is to improve the energy concentration about the IF for monocomponent signals and improve the resolution for multicomponent signals.

Following this logic, we may therefore conclude that, to be a suitable tool for a *practical* time-frequency analysis, a TFD should verify the following minimum set of properties:<sup>4</sup>

1. Be real,
2. Preserve the total energy of the signal:

$$E = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \rho_z(t, f) dt df$$

3. Preserve the regional (component) energy: energy in the region  $R$  of the time-frequency plane bounded in time by  $[t_1, t_2]$  and frequency  $[f_1, f_2]$  should be:

$$E_R = \int_{f_1}^{f_2} \int_{t_1}^{t_2} \rho_z(t, f) dt df$$

4. Reduce the cross-terms, while preserving resolution by minimising measure  $P$  (defined by equation (5)),
5. Reveal the IF law of a monocomponent signal by its peak.

To satisfy these constraints, Barkat and Boashash [4, 5] recently proposed a kernel for a quadratic TFD, known as the B-distribution, defined by:

$$G(t, \tau) = \int_{-\infty}^{\infty} g(\nu, \tau) e^{j2\pi\nu t} d\nu = \left( \frac{|\tau|}{\cosh^2(t)} \right)^\beta \quad (11)$$

The kernel filter  $g(\nu, \tau)$  of the B-distribution (BD) was chosen in the ambiguity domain to be a two-dimensional function centred around the origin with sharp cut-off edges. In this way, the kernel would allow to retain as much auto-terms as possible while filtering out as much cross-terms. The amounts of auto-terms and cross-terms kept and filtered out are functions of the volume underneath the 2-D function  $g(\nu, \tau)$ . This volume can be changed by varying a single parameter  $\beta$  ( $0 \leq \beta \leq 1$ ) which is application dependent.

In addition, a modification to the BD kernel (the Modified B-distribution) by authors Hussain and Boashash allows an efficient estimation of the IF laws of a multicomponent signal.

The kernel of the Modified B-distribution is defined as [10]:

$$G(t, \tau) = \frac{\Gamma(2\alpha)}{2^{2\alpha-1}\Gamma^2(\alpha)} \frac{1}{\cosh^{2\alpha}(t)} \quad (12)$$

where  $\Gamma[\cdot]$  is the gamma function and  $\alpha$  is a real positive number less than 1.

#### 4. PERFORMANCE MEASURE AND COMPARISON OF TIME-FREQUENCY DISTRIBUTIONS

In this section, we use the newly defined measure criterion to compare the performance of the WVD, the spectrogram, the Choi-Williams distribution, the Born-Jordan distribution, Zhao-Atlas-Marks distribution, the B-distribution and the Modified B- (MB) distribution of the two-LFM-component signal defined in Section 1. For each time-frequency distribution we take a slice at the middle of the time interval and measure the parameters  $A_M$ ,  $A_S$ ,  $A_X$  and  $V$ . These parameters are then used to calculate the frequency

<sup>4</sup>Note that the selection of a complete set of properties would be application dependent.

separation of the components  $D$ , defined by equation (4), and the performance indicator  $P$ , defined by equation (5).

The distributions and their respective measurements parameters are recorded in Table 1, while the slices of the TFDs at the middle of the time interval are displayed in Figure 5.

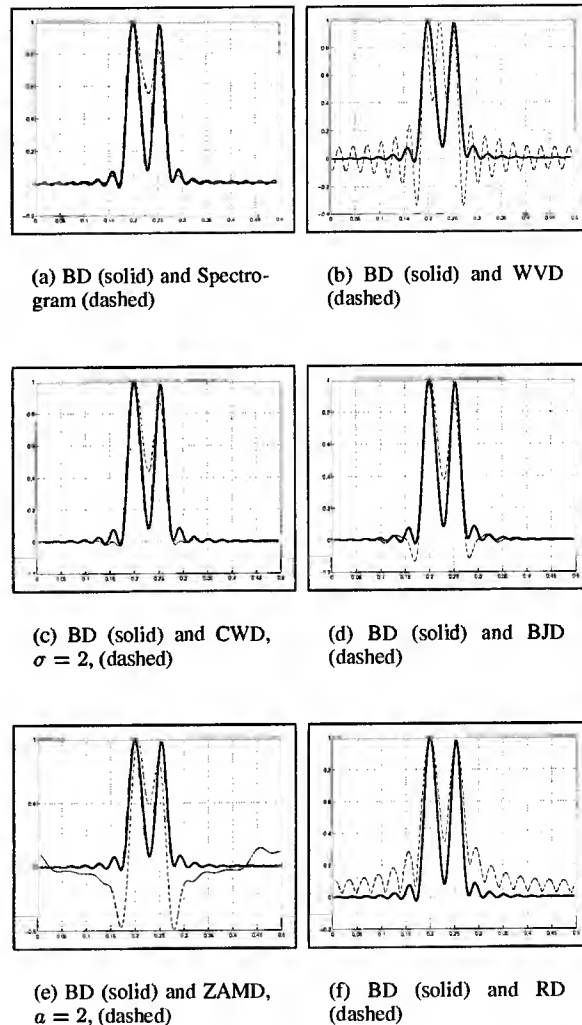


Figure 5: Slices taken at a half of the time interval of TFDs of two closely-spaced LFM signals with frequency  $f_1 = 0.15 - 0.25$  Hz and  $f_2 = 0.2 - 0.3$  Hz. BD=B-distribution, WVD=Wigner-Ville distribution, CWD=Choi-Williams distribution, BJD=Born-Jordan distribution, ZAMD=Zhao-Atlas-Marks Distribution, and RD=Rihaczek Distribution

The TFD which gives the smallest positive  $P$  is the TFD with the best performance when used to analyse multicomponent signals. In our case, the B-distribution ( $\beta = 0.01$ ) yields the smallest value for  $P$  ( $P = 1.04 \times 10^{-2}$ ) and hence is regarded as best. Similar results were obtained with other types of signals.



TFD	$A_M$	$A_S$	$A_X$	$V[Hz]$	$D$	Performance Measure $P$
B-Distribution (BD), $\beta = 0.01$	0.9890	0.0796	0.0810	0.0197	0.6337	$1.04 \times 10^{-2}$
Modified B-Distribution (MBD), $\alpha = 0.01$	0.9885	0.0861	0.0947	0.0199	0.6298	$1.32 \times 10^{-2}$
Born-Jordan Distribution (BJD)	0.9320	0.1227	0.3798	0.0236	0.5164	$1.04 \times 10^{-1}$
Choi-Williams Distribution (CWD), $\sigma = 2$	0.9335	0.0211	0.4415	0.0258	0.4766	$2.25 \times 10^{-2}$
Spectrogram (Hanning window)	0.9119	0.0493	0.5527	0.0323	0.3557	$9.21 \times 10^{-2}$
Rihaczek Distribution (RD)	0.9823	0.2945	0.3446	0.0289	0.4630	$2.71 \times 10^{-1}$
Wigner-Ville Distribution (WVD)	0.9153	0.4134	1	0.0140	0.7558	$6.53 \times 10^{-1}$
Zhao-Atlas-Marks Distribution (ZAMD), $a = 2$	0.9146	0.4822	0.4796	0.0238	0.4331	$6.08 \times 10^{-1}$

Table 1: Measurements parameters and the performance indicator  $P$  of TFDs (slices taken at the half of the signal time interval) of two closely-spaced LFM with frequency  $f_1 = 0.15 - 0.25\text{Hz}$  and  $f_2 = 0.2 - 0.3\text{Hz}$

#### 4.1. Optimisation of the B-Distribution Parameter $\beta$ Using the Performance Measure $P$

The performance measure  $P$  can be used to optimise the value of the smoothing parameters of a given TFD. One approach would be to take consecutive slices of the TFD, find measure  $P$  for each of the slices, and average all such obtained measures for a given value of the TFD parameter to obtain the average performance measure  $P_{av}$ . Repeating this procedure over a range of values of the smoothing parameter, it is possible, by identifying the one which results into smallest  $P_{av}$ , to obtain the optimal value of the smoothing parameter of the TFD considered.

For example, using the measure  $P$ , we can optimise the parameter  $\beta$  of the B-distribution for the signal in Section 1. Simulations have shown that  $\beta = 0.01$  gives visually most appealing results for various multicomponent signals [4]. However, this value can be refined by applying the above described optimisation procedure.

By calculating  $P_{av}$  for  $\beta \in [0, 1]$  with the increment of  $10^{-5}$  and for the distribution slices 16:112 (note that the signal length is  $N = 128$ )<sup>5</sup> we find the optimal value of the smoothing parameter of the B-distribution to be  $\beta_{opt} = 9.9 \times 10^{-4}$  ( $P_{av} = 9.1 \times 10^{-3}$ ). Indeed, a reduction in  $P_{av}$  value of approximately  $2 \times 10^{-3}$  is achieved if the smoothing parameter of the B-distribution is optimised, when compared to  $P_{av} = 1.1 \times 10^{-2}$  of the B-distribution with  $\beta = 0.01$ .

#### 5. CONCLUSION

This paper has presented two key results which we believe to be fundamental to a better understanding and use of time-frequency signal analysis tools.

The first key result is a definition of an objective criterion to compare the resolution performance of time-frequency distributions for multicomponent signals analysis using a quantitative measure of goodness for TFDs. This result fills an obvious need in that until now the comparison of the resolution performance of TFDs was primarily based on a visual impression of the plots of TFDs.

The second key result is an improvement in the design of tools for high resolution time-frequency analysis of multicomponent signals. By removing limitations in the way desirable properties of

quadratic TFDs were previously chosen, a new set of design criteria has been defined. It was found that such defined B-distribution outperforms other existing distributions in terms of time-frequency resolution, as well as cross-terms suppression, when used to represent signals with closely-spaced components in the time-frequency domain.

The combination of these two results is an important breakthrough for the field of time-frequency signal analysis. It opens the way for further research in developing high resolution DSP tools for non-stationary (time-varying) signals by removing unnecessary limitations, and providing a measure of quality of TFDs.

#### 6. REFERENCES

- [1] H. Choi and W. Williams. Improved time-frequency representation of multicomponent signals using exponential kernels. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 37(6):862–871, June 1989.
- [2] B. Boashash. Time-frequency signal analysis. In S. Haykin, editor, *Advances in Spectrum Analysis and Array Processing*, volume 1, chapter 9, pages 418–517. Prentice Hall, 1991.
- [3] Y. Zhao, R. J. Marks, and L. E. Atlas. The use of cone-shaped kernels for generalised time-frequency representations of nonstationary signals. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 38(7):1082 – 1091, July 1990.
- [4] B. Barkat and B. Boashash. High-resolution quadratic time-frequency distribution for multicomponent signals analysis. *IEEE Trans. on Signal Processing*, 1999. Under review.
- [5] V. Sucic, B. Barkat, and B. Boashash. Performance evaluation of the B-distribution. In *Proc. of the Fifth Int. Symposium on Signal Processing and Its Applications, ISSPA 99*, volume 1, pages 267–270, 1999. <http://www.sprc.qut.edu.au/publications/1999/>.
- [6] B. Boashash. Estimating and interpreting the instantaneous frequency of a signal – part 1: Fundamentals; part 2: Algorithms and applications. *Proceedings of the IEEE*, 80(4):519–568, April 1992.
- [7] B. Boashash and V. Sucic. Performance measure of time-frequency distributions. In *CRC Encyclopedia of Signal Processing*, chapter Time-Frequency Analysis. CRC Press, 2001. To appear.
- [8] B. Boashash. Time-frequency signal analysis. In *CRC Encyclopedia of Signal Processing*, chapter Time-Frequency Analysis. CRC Press, 2001. To appear.
- [9] B. Boashash, editor. *Time-Frequency Signal Analysis. Methods and Applications*. Longman Cheshire, 1992.
- [10] Z. Hussain and B. Boashash. High-resolution instantaneous frequency estimation using reduced interference distributions. In *Proceedings of the 10<sup>th</sup> IEEE Workshop on Statistical Signal and Array Processing*, 2000.

<sup>5</sup>We avoid calculations of the measure  $P$  for the first and the last eighth of the TFD slices (i.e. the beginning and the end of the TFD in time) since it is known [2] that in these regions of the time-frequency plane the components resolution is always significantly degraded.



# THE WIGNER DISTRIBUTION FOR ORDINARY LINEAR DIFFERENTIAL EQUATIONS AND WAVE EQUATIONS

*Lorenzo Galleani and Leon Cohen*

City University of New York, 695 Park Ave., New York, NY 10021 USA.

## ABSTRACT

A new method is presented to study systems governed by ordinary linear differential equations and partial differential equations whose solutions are waves. We show that one can obtain a differential equation for the Wigner distribution of the solution of a dynamical equation of evolution. As an example we derive in a new way the equation governing the Wigner distribution for the Schrodinger equation. We also consider differential equations where the forcing terms are random processes.

## 1. INTRODUCTION

Suppose a dynamical variable,  $x(t)$ , is governed by a differential equation, for example by a linear differential equation with constant coefficients,

$$a_n \frac{d^n x}{dt^n} + a_{n-1} \frac{d^{n-1} x}{dt^{n-1}} + \dots + a_1 \frac{dx}{dt} + a_0 x = f(t) \quad (1)$$

where  $f(t)$  is the driving force. Suppose further we want to study the time-frequency properties of the solution by using a bilinear distribution such as the Wigner distribution. The direct way would be to solve for  $x(t)$  and then calculate the Wigner distribution of  $x(t)$ . Our aim is to obtain the differential equation for the Wigner distribution of the solution and hence bypass the necessity for solving Eq. (1). That is if the Wigner distribution (WD) is defined by

$$W(t, \omega) = \frac{1}{2\pi} \int x^*(t - \frac{1}{2}\tau) x(t + \frac{1}{2}\tau) e^{-j\tau\omega} d\tau \quad (2)$$

we want to obtain an equation of motion for  $W(t, \omega)$  directly.

Similarly, suppose we have a wave equation governed by a partial differential equation. We will show that one can obtain an equation for the Wigner distribution of the solution. Of course, in Wigner's original

paper that is what he did for the Schrodinger equation. But we will show how it can be done for any wave equation.

Also, we will show how the methods can be applied to systems with random input.

**Notation.** We define the following Hermitian operators in the space of two dimensional functions of time and frequency,

$$\mathcal{A} = \frac{1}{2j} \frac{\partial}{\partial t} - \omega \quad ; \quad \mathcal{B} = \frac{1}{2j} \frac{\partial}{\partial t} + \omega \quad (3)$$

$$\mathcal{E} = \frac{1}{2j} \frac{\partial}{\partial \omega} + t \quad ; \quad \mathcal{F} = -\frac{1}{2j} \frac{\partial}{\partial \omega} + t \quad (4)$$

It will also be convenient to define the non Hermitian operators,

$$A = j\mathcal{A} = \frac{1}{2} \frac{\partial}{\partial t} - j\omega \quad ; \quad B = j\mathcal{B} = \frac{1}{2} \frac{\partial}{\partial t} + j\omega \quad (5)$$

Differentiation of functions with respect to time will be indicated by

$$\dot{g}(t) = \frac{d}{dt} g(t) \quad ; \quad g^{(n)} = \frac{d^n}{dt^n} g(t) \quad (6)$$

## 2. RELATIONS BETWEEN THE WD OF A SIGNAL AND THE WD OF A MODIFIED SIGNAL

We define the cross Wigner distribution of two time functions,  $x(t)$  and  $y(t)$ , by

$$W_{x,y}(t, \omega) = \frac{1}{2\pi} \int x^*(t - \frac{1}{2}\tau) y(t + \frac{1}{2}\tau) e^{-j\tau\omega} d\tau \quad (7)$$

Now suppose we know  $W_{x,y}(t, \omega)$ , how can one obtain (for example)  $W_{\dot{x},y}(t, \omega)$  and other such quantities. These type of relations are important for reasons that will become apparent in section 3. We have previously obtained these relations [1] and we just list them here and in the appendix we give the derivations. In particular,

$$W_{\dot{x},y} = A W_{x,y} = \left( \frac{1}{2} \frac{\partial}{\partial t} - j\omega \right) W_{x,y} \quad (8)$$

Galleani's permanent address: Dipartimento di Eletttronica, Politecnico di Torino, C.so Duca degli Abruzzi 24, 10129 Torino, Italy.

Work supported by the Office of Naval Research, the NASA JOVE, and the NSA HBCU/MI programs.

$$W_{x,y} = BW_{x,y} = \left( \frac{1}{2} \frac{\partial}{\partial t} + j\omega \right) W_{x,y} \quad (9)$$

Another important relation is the following. Suppose we have the cross Wigner distribution of  $x$  and  $y$  and wish to obtain the cross Wigner distribution of  $f(t)x(t)$  and  $y(t)$  where  $f(t)$  is an arbitrary function. That is, we want to express  $W_{f x, y}(t, \omega)$  in terms of  $W_{x, y}(t, \omega)$ . This is possible, and the result is

$$W_{f x, y}(t, \omega) = f^*(\mathcal{E}) W_{x, y} \quad (10)$$

$$W_{x, f y}(t, \omega) = f(\mathcal{F}) W_{x, y} \quad (11)$$

### 3. LINEAR DIFFERENTIAL EQUATIONS

In our previous paper we considered the case of an ordinary differential equation with constants coefficients. Here we extend the method to the case with time dependent coefficients,

$$a_n(t) \frac{d^n x}{dt^n} + \dots + a_1(t) \frac{dx}{dt} + a_0(t)x = f(t) \quad (12)$$

or using an operator notation

$$\left[ \sum_{k=0}^n a_k(t) D^k \right] x(t) = f(t) \quad (13)$$

As is standard we define the differential operator by  $D = \frac{d}{dt}$ . We now derive associated equations  $W_{x, x}(t, \omega)$  and  $W_{x, f}(t, \omega)$ . We start by evaluating the cross Wigner of equation (13) with respect to  $f(t)$

$$W_{[\sum_k a_k D^k] x, f} = W_{f, f} \quad (14)$$

We have

$$\sum_k W_{[a_k D^k] x, f} = W_{f, f} \quad (15)$$

and applying property (10)

$$\sum_k a_k^*(\mathcal{E}) W_{D^k x, f} = W_{f, f} \quad (16)$$

and using Eq. (8) we have

$$\left[ \sum_k a_k^*(\mathcal{E}) A^k \right] W_{x, f} = W_{f, f} \quad (17)$$

This is the dynamical equation for  $W_{x, f}$ . We now evaluate the cross Wigner of  $x(t)$  with respect to equation (13), and with similar considerations we obtain

$$\left[ \sum_k a_k(\mathcal{F}) B^k \right] W_{x, x} = W_{x, f} \quad (18)$$

To this equation we apply the differential operator that acts on  $W_{x, f}$  in Eq. (17)

$$\begin{aligned} & \left[ \sum_k a_k^*(\mathcal{E}) A^k \right] \left[ \sum_l a_l(\mathcal{F}) B^l \right] W_{x, x} \\ &= \left[ \sum_k a_k^*(\mathcal{E}) A^k \right] W_{x, f} \end{aligned} \quad (19)$$

and we recognize the right hand side to be  $W_{f, f}$ . Hence equation (17) we have

$$\left[ \sum_k a_k^*(\mathcal{E}) A^k \right] \left[ \sum_l a_l(\mathcal{F}) B^l \right] W_{x, x} = W_{f, f} \quad (20)$$

This is the equation of motion for the Wigner of  $x(t)$ .

### 4. POLYNOMIAL COEFFICIENTS

If one considers the ordinary differential equation

$$p_n(t) \frac{d^n x}{dt^n} + \dots + p_1(t) \frac{dx}{dt} + p_0(t)x = f(t) \quad (21)$$

where  $p_0(t), \dots, p_n(t)$  are polynomials in the  $t$  variable, by using the following relations

$$W_{tx, x} = \mathcal{E} W_{x, x} \quad ; \quad W_{x, tx} = \mathcal{F} W_{x, x} \quad (22)$$

and the usual Eqs. (8) and (9), one can readily get the equation for the Wigner distribution for this case. Specializing the general result (20), that is considering  $a_i(t) = p_i(t)$ , for  $i = 0, \dots, n$ , we have that

$$\left[ \sum_k p_k^*(\mathcal{E}) A^k \right] \left[ \sum_l p_l(\mathcal{F}) B^l \right] W_{x, x} = W_{f, f} \quad (23)$$

### 5. CONSTANT COEFFICIENTS

We have previously considered the case of a linear differential equation with constant coefficients [1] and hence we just summarize the results here. We write the equation of motion for  $x(t)$ , as

$$[a_n D^n + a_{n-1} D^{n-1} \dots a_1 D + a_0] x(t) = f(t) \quad (24)$$

and further write it using the standard polynomial notation

$$P_n(D)x(t) = f(t) \quad (25)$$

where

$$P_n(D) = a_n D^n + a_{n-1} D^{n-1} \dots a_1 D + a_0 \quad (26)$$

We have shown that the governing equation for the Wigner distribution is given by,

$$P_n^*(A)P_n(B)W_{x, x} = W_{f, f} \quad (27)$$

and we have also shown that solving this equation directly has significant advantages [2].

## 6. WAVE EQUATIONS

The same approach that we adopted in section 3 for ordinary differential equations, can be generalized to linear partial differential equations, and in particular to wave equations. Wigner in his original paper obtained the equation of evolution for the Wigner distribution for the Schrodinger equation. It is our aim to develop the Wigner distribution approach for *classical* wave equations, such as electromagnetic or acoustic wave equations. Here we will discuss some issues in regard to this approach and a fuller development will be given in a future publication. Suppose we have a partial differential equation for  $u(x, t)$ , for example, wave equation. As Wigner did for the Schrodinger wave function we can define the Wigner distribution of position, momentum, and time by

$$W_{\psi, \psi}(x, p, t) = \frac{1}{2\pi} \int \psi^*(x - \frac{1}{2}\tau_x, t) \psi(x + \frac{1}{2}\tau_x, t) \times e^{-j\tau_x p} d\tau_x \quad (28)$$

This is the approach that Wigner took in his original paper. In this approach time  $t$  plays a passive role. However one can also define a joint distribution of the four variables<sup>1</sup>

$$K_{\psi, \psi}(x, p, t, \omega) = \frac{1}{(2\pi)^2} \int \psi^*(t - \frac{1}{2}\tau, x - \frac{1}{2}\tau_x) \times \psi(t + \frac{1}{2}\tau, x + \frac{1}{2}\tau_x) e^{-j\tau\omega - j\tau_x p} d\tau d\tau_x \quad (29)$$

It follows that

$$\int K_{\psi, \psi}(x, p, t, \omega) d\omega = W_{\psi, \psi}(x, p, t) \quad (30)$$

Now, a fundamental issue is whether one can obtain equations of evolution for  $W$  and  $K$ . We believe that one may always obtain an equation for  $K$  but not always for  $W$ ! In another paper we will discuss this issue in detail but here we illustrate our method by examining two wave equations, the Schrodinger equation and the classical wave equation. We emphasize that our aim in devising these methods is to study classical equations of motions by way of the Wigner distribution.

### 6.1. Operator Relations

It is possible to prove the following results

$$W_{\frac{\partial \psi}{\partial x}, \psi} = A_x W_{\psi, \psi} ; \quad W_{\psi, \frac{\partial \psi}{\partial x}} = B_x W_{\psi, \psi} \quad (31)$$

where

$$A_x = \frac{1}{2} \frac{\partial}{\partial x} - jp ; \quad B_x = \frac{1}{2} \frac{\partial}{\partial x} + jp \quad (32)$$

<sup>1</sup>We use  $K_{\psi, \psi}(x, p, t, \omega)$  for notational clarity, that is to contrast with  $W_{\psi, \psi}(x, p, t)$ .

and also equations

$$W_{f\psi, \psi} = f^*(\mathcal{E}_x) W_{\psi, \psi} ; \quad W_{\psi, g\psi} = g(\mathcal{F}_x) W_{\psi, \psi} \quad (33)$$

where

$$\mathcal{E}_x = \frac{1}{2j} \frac{\partial}{\partial p} + x ; \quad \mathcal{F}_x = -\frac{1}{2j} \frac{\partial}{\partial p} + x \quad (34)$$

and where  $f = f(x)$ ,  $g = g(x)$  are two arbitrary functions.

### 6.2. The Schrodinger Equation

The Schrodinger Equation is a wave equation for which an equation of motion for both  $W$  and  $K$  can be obtained. We first obtain the equation for  $W$ . Schrodinger Equation is<sup>2</sup>

$$j \frac{\partial \psi(x, t)}{\partial t} = -\frac{1}{2m} \frac{\partial^2 \psi(x, t)}{\partial x^2} + V(x) \psi(x, t) \quad (35)$$

To obtain the equation for the Wigner distribution  $W_{\psi, \psi}$ , we evaluate the cross Wigner of Eq. (35) with respect to  $\psi(x, t)$

$$W_{j \frac{\partial \psi}{\partial t}, \psi} = W_{-\frac{1}{2m} \frac{\partial^2 \psi}{\partial x^2}, \psi} + W_{V\psi, \psi} \quad (36)$$

Extracting the coefficients we have

$$-j W_{\frac{\partial \psi}{\partial t}, \psi} = -\frac{1}{2m} W_{\frac{\partial^2 \psi}{\partial x^2}, \psi} + W_{V\psi, \psi} \quad (37)$$

Applying properties (31) and (33) we obtain

$$-j W_{\frac{\partial \psi}{\partial t}, \psi} = -\frac{1}{2m} A_x^2 W_{\psi, \psi} + V^*(\mathcal{E}_x) W_{\psi, \psi} \quad (38)$$

Now we evaluate the cross Wigner of  $\psi(x, t)$  with respect to Eq. (35)

$$W_{\psi, j \frac{\partial \psi}{\partial t}} = W_{\psi, -\frac{1}{2m} \frac{\partial^2 \psi}{\partial x^2}} + W_{\psi, V\psi} \quad (39)$$

and with similar operations we get

$$j W_{\psi, \frac{\partial \psi}{\partial t}} = -\frac{1}{2m} B_x^2 W_{\psi, \psi} + V(\mathcal{F}_x) W_{\psi, \psi} \quad (40)$$

We then subtract Eq. (40) from Eq. (38), getting

$$-j \left[ W_{\frac{\partial \psi}{\partial t}, \psi} + W_{\psi, \frac{\partial \psi}{\partial t}} \right] = -\frac{1}{2m} [A_x^2 - B_x^2] W_{\psi, \psi} + [V^*(\mathcal{E}_x) - V(\mathcal{F}_x)] W_{\psi, \psi} \quad (41)$$

<sup>2</sup>We take  $\hbar = 1$ .

Using the fact that

$$\frac{\partial W_{\psi,\psi}}{\partial t} = W_{\frac{\partial \psi}{\partial t},\psi} + W_{\psi,\frac{\partial \psi}{\partial t}} \quad (42)$$

and

$$A_x^2 - B_x^2 = -2jp \frac{\partial}{\partial x} \quad (43)$$

we obtain

$$-j \frac{\partial W_{\psi,\psi}}{\partial t} = j \frac{p}{m} \frac{\partial W_{\psi,\psi}}{\partial x} + [V^*(\mathcal{E}_x) - V(\mathcal{F}_x)] W_{\psi,\psi} \quad (44)$$

One can show that

$$V^*(\mathcal{E}_x) W_{\psi,\psi}(x, p, t) = \frac{1}{2\pi} \iint V^*(x + u/2) e^{-j(p'-p)u} W_{\psi,\psi}(x, p') du dp' \quad (45)$$

$$V(\mathcal{F}_x) W_{\psi,\psi}(x, p, t) = \frac{1}{2\pi} \iint V(x + u/2) e^{j(p'-p)u} W_{\psi,\psi}(x, p') du dp' \quad (46)$$

Now consider real potentials, then

$$[V(\mathcal{E}_x) - V(\mathcal{F}_x)] W_{\psi,\psi} = -\frac{j}{\pi} \int V(x + u/2) \sin[(p' - p)u] W_{\psi,\psi}(x, p') du dp' \quad (47)$$

and hence we have that

$$\frac{\partial W_{\psi,\psi}}{\partial t} + j \frac{p}{m} \frac{\partial W_{\psi,\psi}}{\partial x} = \frac{1}{\pi} \int V(x + u/2) \sin[(p' - p)u] W_{\psi,\psi}(x, p') du dp' \quad (48)$$

Which is the well known result for the Wigner distribution and was derived by Wigner using different methods [6].

We now derive an equation for  $K$ . First we point out that Eqs. (10) and (11) still hold, and that the following relationships can be easily proved with the same technique of the ordinary equation case

$$K_{\frac{\partial \psi}{\partial x},\psi} = A_x K_{\psi,\psi} \quad K_{\frac{\partial \psi}{\partial t},\psi} = A_t K_{\psi,\psi} \quad (49)$$

$$K_{\psi,\frac{\partial \psi}{\partial x}} = B_x K_{\psi,\psi} \quad K_{\psi,\frac{\partial \psi}{\partial t}} = B_t K_{\psi,\psi} \quad (50)$$

where

$$A_x = \frac{1}{2} \frac{\partial}{\partial x} - jp \quad A_t = \frac{1}{2} \frac{\partial}{\partial t} - j\omega \quad (51)$$

$$B_x = \frac{1}{2} \frac{\partial}{\partial x} + jp \quad B_t = \frac{1}{2} \frac{\partial}{\partial t} + j\omega \quad (52)$$

and  $K_{\psi,\psi} = K_{\psi,\psi}(x, p, t, \omega)$  is the four dimensional Wigner distribution. Using the same approach of section 6.2, we evaluate the cross Wigner of the two sides of Schrodinger's equation with respect to  $\psi$ , and applying the new operator relations, we have

$$-j A_t K_{\psi,\psi} = -\frac{1}{2m} A_x^2 K_{\psi,\psi} + V^*(\mathcal{E}_x) K_{\psi,\psi} \quad (53)$$

Now we do the other way, taking the cross Wigner of  $\psi$  with respect to Schrodinger's equation, obtaining

$$j B_t K_{\psi,\psi} = -\frac{1}{2m} B_x^2 K_{\psi,\psi} + V(\mathcal{F}_x) K_{\psi,\psi} \quad (54)$$

We subtract Eq. (54) from Eq. (53), and we have

$$-j [A_t + B_t] = -\frac{1}{2m} [A_x^2 - B_x^2] K_{\psi,\psi} + [V^*(\mathcal{E}_x) - V(\mathcal{F}_x)] K_{\psi,\psi} \quad (55)$$

Using the fact that  $A_t + B_t = \frac{\partial}{\partial t}$  and  $A_x^2 - B_x^2 = -2jp \frac{\partial}{\partial x}$  we obtain

$$-j \frac{\partial K_{\psi,\psi}}{\partial t} = j \frac{p}{m} \frac{\partial K_{\psi,\psi}}{\partial x} + [V^*(\mathcal{E}_x) - V(\mathcal{F}_x)] K_{\psi,\psi} \quad (56)$$

We emphasize that while this equation looks the same as Eq. (44) it is not because  $K$  is a four dimensional density. One can obtain Eq. (44) by integrating out  $\omega$ . This is due to the fact that all the operators that act on  $K_{\psi,\psi}$  do not involve  $\omega$  and hence one can obtain the equation of motion for  $W$  Eq. (44) from the equation of motion of  $K$ .

## 7. CLASSICAL WAVE EQUATION

We want to apply the same method developed for the case of ordinary differential equations to the classic wave equation

$$\left[ \frac{\partial^2}{\partial x^2} - \frac{1}{c^2} \frac{\partial^2}{\partial t^2} \right] u(x, t) = f(x, t) \quad (57)$$

One can prove with the same considerations of section 3 that the equation for the four dimensional Wigner distribution is

$$\left[ A_x^2 - \frac{1}{c^2} A_t^2 \right] \left[ B_x^2 - \frac{1}{c^2} B_t^2 \right] K_{\psi,\psi}(x, p, t, \omega) = K_{f,f}(x, p, t, \omega) \quad (58)$$

However we believe that it is impossible to obtain an equation for the three dimensional Wigner distribution in this case. One can convince oneself of this by attempting to do so directly by the same methods that have been applied to the Schrodinger equation. Alternatively one attempt to get it is by integrating out  $\omega$  from Eq. (58). But it is not possible to integrate out  $\omega$  to obtain an equation for  $W(x, p, t)$ . This is due to the fact that the operators  $A_t$  and  $B_t$  contain  $\omega$ .

## 8. RANDOM DRIVING FORCE AND RANDOM COEFFICIENTS

There are two important cases that arise in many areas for physics, chemistry and engineering. The first case is where only the random force is stochastic. That is the most important case. The other case is where the coefficients are randomly given. In this paper, for lack of space, we only consider the first case with deterministic coefficients, that is Eq. (12) with a random driving force. To illustrate we consider here the case of an harmonic oscillator with a Gaussian process  $N(t)$  as input

$$\ddot{x}(t) + 2\mu\dot{x}(t) + \omega_0^2 x(t) = N(t) \quad (59)$$

The governing equation for the Wigner distribution of  $x(t)$  is, from (20)

$$[A^2 + 2\mu A + \omega_0^2] [B^2 + 2\mu B + \omega_0^2] W_{x,x} = W_{N,N} \quad (60)$$

Here obviously both  $x(t)$  and  $W_{x,x}(t, \omega)$  are stochastic processes. The important fact of having a stochastic equation for the Wigner distribution, is that one can get deterministic equations for the mean values, and for any moment in general. For example, considering that a linear (differential) operator acts on  $W_{x,x}$  in Eq. (60), one can evaluate the ensemble average of the two sides of the same equation

$$[A^2 + 2\mu A + \omega_0^2] [B^2 + 2\mu B + \omega_0^2] \mathcal{E}[W_{x,x}] = \mathcal{E}[W_{N,N}] \quad (61)$$

More interesting is the case of an harmonic oscillator with time-varying coefficients. For example, if we consider the underdamped case with  $\mu \ll \omega_0$ , then it is known that the harmonic oscillator behaves like a bandpass filter with central frequency

$$\omega_c = \sqrt{\omega_0^2 - \mu^2} \approx \omega_0 \quad (62)$$

By letting  $\omega_0 = \omega_0(t)$  we can hence build a bandpass filter with time-varying central frequency. If again we set a Gaussian process as input we have

$$\ddot{x}(t) + 2\mu\dot{x}(t) + \omega_0^2(t)x(t) = N(t) \quad (63)$$

This equation is of interest in many areas, since many systems show a bandpass behavior which is actually time or space varying under closer analysis. The differential equation for the Wigner distribution is, in this case,

$$[A^2 + 2\mu A + \omega_0^2(\mathcal{E})] [B^2 + 2\mu B + \omega_0^2(\mathcal{F})] W_{x,x} = W_{N,N} \quad (64)$$

The interesting thing is that we have a "stationary" process  $N(t)$  as input that is processed by a time-varying system that generates a "nonstationary" process  $x(t)$  as output. Now, an important aspect of such a process is the random instantaneous frequencies in  $x(t)$ , which is due to the time-varying filtering. It is hence important to derive the differential equation (64) for the Wigner  $W_{x,x}(t, \omega)$  of  $x(t)$ , because from it we may be able to derive equations for moments of  $W_{x,x}(t, \omega)$  which are physically significant. For example, taking the ensemble average of both sides of Eq. (64), we have

$$[A^2 + 2\mu A + \omega_0^2(\mathcal{E})] [B^2 + 2\mu B + \omega_0^2(\mathcal{F})] \mathcal{E}[W_{x,x}] = \mathcal{E}[W_{N,N}] \quad (65)$$

This approach may be fruitful for studying how the moments of the input process evolve in time.

## 9. CONCLUSION

We have derived a method to study systems governed by linear ordinary and partial differential equations. The method allows us to write an associate equation for the Wigner distribution of the solution. In this paper we have shown how to transform ordinary equations with time-varying coefficients. We have also given examples on how the method can be used to study wave equations and systems with random inputs.

## REFERENCES

- [1] L. Galleani, L. Cohen, "Dynamics Using the Wigner Distribution", to appear in ICPR 2000, September 3-8 2000, Barcelona, Spain.
- [2] L. Galleani, L. Cohen, "On the Exact Solution to the "Gliding Tone" Problem", to appear in IEEE SSAP 2000, August 14-16 2000, Pocono Manor, Pennsylvania, USA
- [3] L. Cohen, *Time-Frequency Analysis*, Prentice-Hall, 1995.
- [4] W. T. Thomson, *Theory of Vibrations with Applications*, Chapman and Hall, 1983.
- [5] N. Wax, *Selected papers on random noise and stochastic processes*, Dover Publications, Inc.
- [6] E. P. Wigner, "On the quantum correction for thermodynamic equilibrium," *Physical Review*, vol. 40, pp. 749-759, 1932.

# APPLICATION OF TIME-FREQUENCY TECHNIQUES FOR THE DETECTION OF ANTI-PERSONNEL LANDMINES

*B. Barkat, A.M. Zoubir and C.L. Brown*

Communications & Signal Processing Group  
Australian Telecommunications Research Institute  
School of Electrical & Computer Engineering  
Curtin University of Technology  
GPO Box U1987, Perth, WA 6845, Australia  
Tel: +61-8-9266-7890; Fax: +61-8-9266-2584  
E-mail: rbarkatb@curtin.edu.au

## ABSTRACT

In this paper we propose two methods to detect buried underground objects. Both methods are based on time-frequency analysis. The first approach uses the instantaneous frequency of the return Ground Penetrating Radar (GPR) signals and the second approach uses a time-frequency distribution, such as the Wigner-Ville or the spectrogram, of the return signals. Real data were used in the examples to validate the proposed algorithms.

## 1. INTRODUCTION

Landmines are causing enormous humanitarian and economic problems in many countries around the world. Experts estimate that up to 110 millions of landmines are still to be cleared and that more than 500 civilians are killed or maimed every week by landmines. Most of the victims are innocent children [1].

Today, the most widespread technique for landmine detection is metal detection [3]. However, this technique becomes almost useless when there is a large amount of metal debris in the field to be cleared. In this case, manual probing is required and the demining process becomes very laborious and slow. In addition, modern war technologies are producing mines that contain no or a very small amount of metal.

To avoid the above mentioned problems Ground Penetrating Radar (GPR) has been applied. The use of GPR systems stems from their ability to detect buried objects based on a change in the dielectric permittivity of the ground rather than the metal content of the target [5].

Extensive data analysis shows that the GPR return signal is non-stationary. Thus, by using time-frequency

techniques in the analysis of the return signal we may be able to detect a target more accurately. The signal feature selected for detection is the instantaneous power of the return signal and/or its energy evaluated using a time-frequency distribution, namely, the Wigner-Ville distribution (WVD).

## 2. PRELIMINARIES

The WVD is defined as

$$W(t, f) = \int_{-\infty}^{+\infty} z(t + \frac{\tau}{2}) \cdot z^*(t - \frac{\tau}{2}) e^{-j2\pi f\tau} d\tau \quad (1)$$

where  $z(t)$  is the analytic signal associated with the real signal under consideration,  $s(t)$  [4].

We can show that integrating the WVD,  $W(t, f)$ , over all frequencies would result in the instantaneous signal power,  $|z(t)|^2$ ; while its integration over time would result in the energy spectrum  $|Z(f)|^2$ . We can also obtain the total signal energy  $E_z$  by integrating the WVD over time and frequency as follows

$$E_z = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} W(t, f) df dt \quad (2)$$

An important concept closely related to the time-frequency analysis is the instantaneous frequency (IF) defined as

$$f_i(t) = \frac{1}{2\pi} \frac{d\phi(t)}{dt} \quad (3)$$

where  $\phi(t)$  is the phase of the signal under consideration.

The above definition and results will play a major role in the detection procedure outlined below.

### 3. GPR RETURN SIGNALS

In this section we give a brief description of the experiment and outline the need for a further processing of the GPR data.

The experiment was conducted at the Defence Science & Technology Organisation (DSTO) Australia using a GPR equipped with a bistatic bow-tie antenna. The antenna is moved, in one direction over a distance  $d$ , above the ground surface at a constant velocity and constant height. At regular time instants, the GPR system radiates short duration pulses of electromagnetic energy into the ground and collects backscattered signals. A target is declared present if the GPR detects a local change (or discontinuity) in the soil dielectric.

The collection of the regularly spaced return signals, referred as GPR traces, over the distance  $d$  is called the radargram. In Figure 1, we display a typical radargram obtained from an experiment discussed in detail in the next sections.

In the experiment we buried three different targets; however, the radargram reveals the potential existence of two targets only and misses the third target.

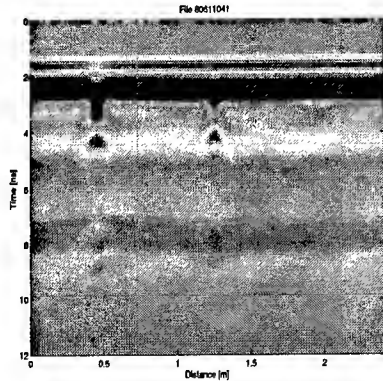


Figure 1: Radargram of three landmines buried in a dry sandpit (file 80611041).

The time domain plots of two different GPR traces, one taken at a position where a target exists the other taken at a position where there is no target, are much alike and do not give specific answer about the presence of a target (see Figure 2).

In order to improve the detection performance other techniques have been suggested [3, 6]. Here, we propose two different techniques: an IF based detection and an energy based detection.

### 4. THE IF BASED TECHNIQUE

The ability of the time-frequency distribution to display the signal's spectral components makes it a very

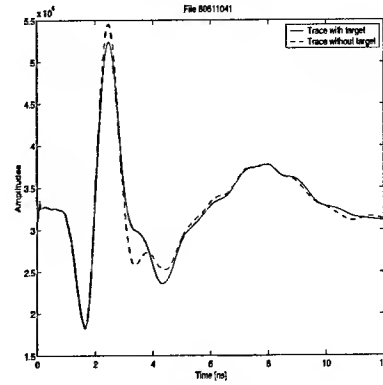


Figure 2: Two GPR return signals: one with target, the other without target.

powerful tool in the localisation and estimation of the instantaneous frequency of the signal.

For the estimation of the IF of the GPR signals we use the peak of the WVD. The change in the IF from one trace to another cannot be seen directly from the time-frequency representations, of the traces, nor from the plots of the IF estimates (not shown here). However, by defining a measure criterion as follows

$$T_k = \int_{-\infty}^{\infty} [f_i^k(t) - f_i^{back}(t)]^2 dt \quad k = 1, \dots, N_s, \quad (4)$$

we are able to detect the presence of the target. In Equation (4)  $f_i^k(t)$  refers to the IF of the  $k^{th}$  GPR trace under analysis,  $f_i^{back}(t)$  refers to the IF of a background only (no target present) and  $N_s$  is the total number of GPR traces in the radargram.

We should note that prior to estimating the IF, we subtract the mean of the signal and form its analytic version in order to avoid aliasing in the time-frequency distribution [2]. Thus, the IF based detection algorithm can be stated as given by Table 1 below.

As an example, consider the detection of two surrogate anti-personnel landmines, referred as ST-AP(2) and ST-AP(3), modeled after the PMN and PMN2 respectively and buried in a dry sandpit. These two targets have no metal in their casings and have dimensions of 11.8cm×5.0cm ( $\phi \times h$ ) and 11.5cm×5.3cm respectively. The first target is located at 71.9 cm from the origin while the second target is located at 156.1 cm from the origin. The radargram of this experiment is shown in Figure 3. Signals taken from the radargram at two different positions (one where a target is known to be present the other where there is no target) are displayed in Figure 4. It is clear from the two plots that neither the radargram nor the time domain can discriminate between the target and the background (target-free). By using the proposed IF based algorithm, we obtain the

For every GPR trace of the radargram  $s_k(t)$   $1 \leq k \leq N_s$ ,

1. Remove the mean of the signal

$$x_k(t) = s_k(t) - \text{mean}(s_k(t))$$

2. Compute the analytic version of  $x_k(t)$

$$z_k(t) = x_k(t) + j\mathcal{H}[x_k(t)]$$

where  $\mathcal{H}[\cdot]$  is the Hilbert transform operation.

3. Compute the time-frequency distribution  $W_z(t, f)$  of the signal  $z_k(t)$ . The IF estimates are found as the maximum of  $W_z(t, f)$  for each time instant  $t$ , i.e.,

$$f_i^k(t) = \arg[\max_I W_z(t, f)]$$

where  $I = \{f : 0 \leq |f| \leq 1/2T\}$  and  $T$  being the sampling period of the signal.

4. Compute the measure criterion as given by (4).

Table 1: IF based detection algorithm

results displayed in Figure 5. We can clearly observe the presence of the targets at their respective positions.

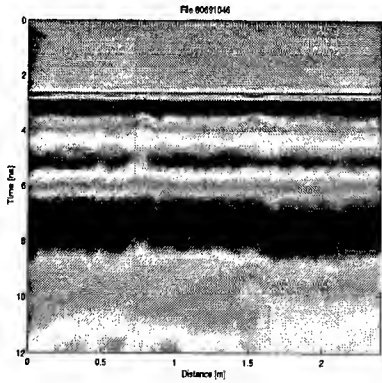


Figure 3: Radargram of two surrogate anti-personnel landmines.

Note that the phases of the GPR signals computed at all positions  $d$  can also be used to locate the target over the analysed distance. However, for some situations the algorithm does not detect all the targets present as is the case for the above example. Figure 6 illustrates the point.

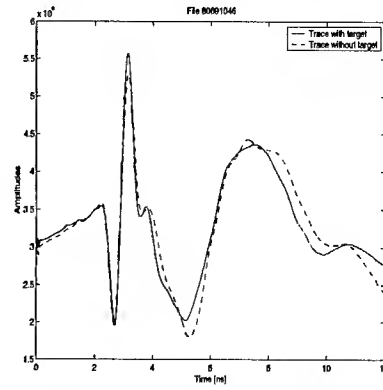


Figure 4: Two GPR return signals: one with target, the other without target.

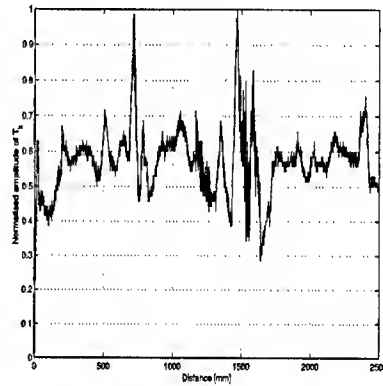


Figure 5: IF based algorithm result for the detection of two surrogate anti-personnel landmines.

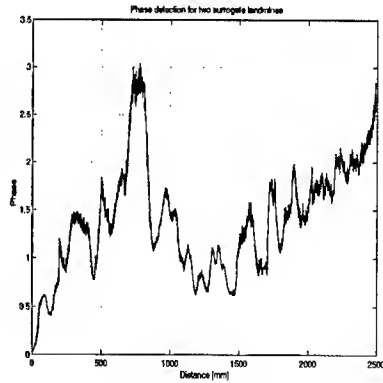


Figure 6: Phase based algorithm result for the detection of two surrogate anti-personnel landmines.

## 5. ENERGY BASED DETECTION

In this section, we use another time-frequency approach to detect the presence of a buried target in the soil. The



method is based on the discriminator  $D$  defined as

$$D = \int \int W_{(x-back)}(t, f) dt df \quad (5)$$

where  $W_z(t, f)$  represents the WVD of a given signal  $z$ . In the above equation,  $W_{(x-back)}(t, f)$  represents the WVD of the difference of two signals:  $x$ , the actual return signal under analysis and  $back$ , a reference background signal (no target). Here again, the aim is to decide whether the actual signal  $x$  represents a GPR trace where a target exists or a GPR trace where there is no target. We should note that other time-frequency distributions, such as the spectrogram, can also be used in place of the WVD.

The algorithm has been applied to a large number of radargrams with different types of targets and different types of soil. The results show that the method can effectively reveal the presence of a buried target. As an illustration, let us consider the detection of three surrogate landmines buried in sandpit at depth 0.5 cm and at distances 0.565m, 1.537m and 2.413m respectively. Two of these targets are made of solid stainless steel cylinders and have dimensions of 10cm  $\times$  5cm ( $\phi \times h$ ) and 5cm  $\times$  5cm; whereas, the third target is a PVC cylinder and has dimensions 10cm  $\times$  5cm. The radargram of this experiment is displayed in Figure 1 and the time representation of two traces (with and without target) are displayed in Figure 2. As stated earlier it is very difficult to detect all the targets from these two plots. However, when we apply the algorithm based on the discriminator  $D$  we obtain the results displayed in Figure 7.

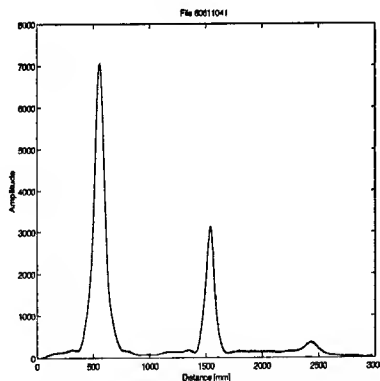


Figure 7: Discriminant based algorithm result for the detection of three buried targets in sandpit.

Note that the energy return of the bigger stainless steel cylinder is much higher than the energy return of the smaller steel cylinder or the PVC cylinder.

Before concluding, we should emphasise that the two techniques proposed here are able to detect buried

objects in the ground including unwanted targets, such as scrap metal. In order to detect only wanted targets, a thresholding and a classification procedure has to follow the detection procedure. The classification is beyond the scope of the present paper and will be addressed elsewhere.

## 6. CONCLUSION

In this paper, we proposed two different methods to detect buried objects underground. Both methods are based on time-frequency techniques. The first method uses the instantaneous frequency of the return GPR signal; whereas, the second approach uses a discriminant measure evaluated using the time-frequency distribution of the signal. Examples, using real data, show that both algorithms are very powerful in detecting buried landmines.

## Acknowledgement

The authors would like to thank Dr. Ian Chant for his technical assistance and for providing the data used in the examples.

## 7. REFERENCES

- [1] Adopt-a-minefield. World Wide Web, February 2000. <http://www.landmines.org>.
- [2] B. Boashash. Time-Frequency Signal Analysis. In Simon Haykin, editor, *Advances in Spectrum Analysis and Array Processing*, volume 1, chapter 9, pages 418–517. Prentice-Hall, NJ, USA, 1991.
- [3] Hakan Brunzell. *Signal Processing Techniques for Detection of Buried Landmines using Ground Penetrating Radar*. PhD thesis, ECE, Chalmers University of Technology, Sweden, 1998.
- [4] L. Cohen. *Time-Frequency Analysis*. Prentice-Hall, 1995.
- [5] A. M. Zoubir, D. R. Iskander, I. Chant, and D. Carevic. Detection of Landmines Using Ground-Penetrating Radar. In *Proc. SPIE: Detection and Remediation Technologies for Mines and Minelike Targets IV*, volume 3710, pages 1301–1312, Orlando, USA, August 1999.
- [6] A.M. Zoubir and D.R. Iskander. Application of Bootstrap Methods to the Detection of Landmines. Technical Report 3/98-99, Queensland University of Technology, Australia, 1999.

# A NEW MATRIX DECOMPOSITION BASED ON OPTIMUM TRANSFORMATION OF THE SINGULAR VALUE DECOMPOSITION BASIS SETS YIELDS PRINCIPAL FEATURES OF TIME-FREQUENCY DISTRIBUTIONS

Dale Groutage<sup>1</sup>, *Senior Member, IEEE* ;

Naval Surface Warfare Center (NSWC)<sup>3</sup> ;  
Puget Sound Detachment  
530 Farragut Avenue  
Bremerton, WA 98314-5215

Telephone: (360) 476-5927 ;

Fax: (360) 476-5281 ;

David Bennink<sup>2</sup>

Applied Measurements Systems Intl.  
Division of ManTech  
Bremerton, WA 98337

Telephone: (360) 479-5517

Fax: (360) 479-5592

**Abstract**—The classification of objects or quantities in all fields of science depends on the quality of the features used for classifying them. This includes, for example, classification of phenomenon described by nonstationary processes such as electrocardiograms, seismic geophysics signals, submarine transient acoustic signals, and speech signals for recognition. This paper presents a new matrix decomposition that is used to obtain a set of principal features from time-frequency representations for classifying nonstationary time series processes. This new matrix decomposition is based a transformation of the orthonormal basis from singular value decomposition (SVD). The new basis set yields extrema of the first moment for each vector in the new basis set. These basis sets for time and frequency can then be used to construct features relating to the location and spread of each energy density highlight in the time-frequency plane. This new matrix decomposition is presented in this paper along with a simple example to illustrate its application.

## I. Introduction

The development of modern techniques to process nonstationary signals has, and continues to be, the

focus of much research where a goal is the description of the distribution of signals energy as a joint function of time and frequency. To this end, some notable contribution have been made by Wigner [1-2], Cohen [3-5], Choi and Williams [6], Zhao Atlas and Marks [7], and Loughlin, Pitton and Atlas [8], Groutage [9], and Loughlin [10] to name but a few. It is the representation of nonstationary processes by time-frequency distributions that make possible the means for classifying such processes. Certainly, from a classification standpoint, a desirable time-frequency representation is one that has the correct description of energy density. For if this is true, it is unique. Unfortunately, this is not always the case. In fact, it most likely is never the case. For example, the Wigner distribution always satisfies the frequency and time marginals, but is not always manifestly positive and may contain cross products that do not relate to the physical quantities. On the other hand, the spectrogram never satisfies the time and frequency marginals, but is always positive. The spectrogram is a window based time-frequency representation and therefore, is not unique. For a time-frequency distribution to be interpreted as a joint time-frequency energy density, it must satisfy the two fundamental properties of

<sup>1</sup> Dale Groutage is with the Naval Surface Warfare Center (NSWC), Carderock Division, Puget Sound Detachment, Bremerton, WA 98314-5215.

<sup>2</sup> David Bennink is with Applied Measurement Systems Intl., a Division of ManTech, Bremerton, WA 98337

<sup>3</sup> This work was jointly sponsored by the Office of Naval Research (ONR) under direction of Dr. John Tague and Project leader Stephen Greineder at The Naval Under Sea Warfare Center (NUWC) and The Naval Surface Warfare Center (NSWC) In-House Laboratory Independent Research (ILIR) Program under the direction of Dr. John Barkyoub

nonnegativity and the correct time and frequency marginals:

$$Q(t, f) \geq 0 \quad (1)$$

$$\int_{-\infty}^{\infty} Q(t, f) dt = |S(f)|^2 \quad (2a)$$

$$\int_{-\infty}^{\infty} Q(t, f) df = |s(t)|^2 \quad (2b)$$

Where  $S(f) = \int_{-\infty}^{\infty} s(t) e^{-j2\pi ft} dt$  is the Fourier

transform of the signal. The joint time-frequency energy density is a physical quantity that can be used to describe the behavior of the process. It specifies where, jointly in time and frequency, the energy is concentrated. The time marginal by itself describes the instantaneous energy and specifies where in time the energy is located. The frequency marginal in contrast specifies where in frequency the energy is concentrated. However, the joint time-frequency description of energy concentration provides a means for describing the main attributes of the nonstationary process with a relatively few descriptors. It is the essence of these descriptors that can be used for classifying the underlying process.

Marinovic and Eichmann [11] and [12] looked at a feature extraction technique based on the singular value decomposition (SVD) of the Wigner distribution. Their technique used only the singular values to determine the features. In contrast, Groutage and Bennink [13] looked at a new method that uses not only the singular values, but also the singular vectors. The reason being, the singular values are pure numbers and do not contain significant information about the underlying process, whereas, the singular vectors contain the bulk of the information. Since the SVD singular vectors are orthonormal, the vectors whose elements are composed of the squared-elements of the SVD vectors are discrete density functions. Moments generated from these density functions are the principal features of the non-stationary time series process. When the energy density is not uniformly concentrated at various locations in the time frequency plane, this technique works relatively well. However, when the energy is uniformly concentrated at more than one location,

the technique breaks down. It was for this reason that a new matrix decomposition was looked at.

## II. Principal Features From Singular Value Decomposition (SVD)

A matrix  $\mathbf{A}$  can be decomposed into a sum over a set of basis matrices  $\mathbf{A}_i$  each multiplied by a weight  $\sigma_i$ :

$$\mathbf{A} = \sum_{i=1}^R \sigma_i \mathbf{A}_i \quad (3)$$

Although this can be accomplished in a variety of ways, one convenient approach is given by the singular value decomposition Lawson [14]. The SVD yields as the weights a set of positive real numbers, the singular values, such that  $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_R > 0$ , and associated singular vectors  $\mathbf{u}_i$  and  $\mathbf{v}_i$  such that

$$\mathbf{A}_i = \mathbf{u}_i \mathbf{v}_i^H \quad (4)$$

where  $R$  is the rank of  $\mathbf{A}$  and  $H$  is the Hermitian transpose. All of the information contained in  $\mathbf{A}$  is certainly also contained in the complete set of basis matrices  $\mathbf{A}_i$  and weights  $\sigma_i$ . While the complete form of (3) does not provide a reduction to a small set of descriptors, the hope is that the decomposition method leads to an easier way to extract the important information. This is certainly the case if the weights alone can be used as the derived features.

If the basis matrices  $\mathbf{A}_i$  were a fixed set, then the index on the singular values  $\sigma_i$  would associate directly with time-frequency content, since the time-frequency distribution would be known for each  $\mathbf{A}_i$ . However, the basis set is determined as an integral part of the SVD. Thus, in order to assign time-frequency content to each  $\sigma_i$  it is necessary to extract this information from the  $\mathbf{A}_i$ , which together with the singular values will provide the desired features. A naturally first step is to use the joint time-frequency moments in (3) to characterize the  $\mathbf{A}_i$ . Under the assumption that the SVD process has separated the energy highlights, only a few such moments should be required. An added benefit of the  $\mathbf{A}_i$  from (3) is that the time and frequency aspects are independent, so only the temporal and spectral moments need to be considered, as opposed to joint moments.

## III. Principal Features From Transformed Singular Value Decomposition (TSVD)

When the above assumption is not met, i.e., SVD process cannot separate the energy highlights, the

resulting features will associate with a linear combinations of a number of the singular vectors in time and frequency. These resulting features would not, for the most part, be useful for classification purposes. However, by studying a few simple examples that pointed out the dilemma of the energy density highlights being associated by linear combinations of respective singular time or frequency vectors, the idea was posed to construct a new basis set. Basically, the idea is to rotate the original SVD basis vectors to a new orientation in the span of their vector space such as to minimize the number of vectors required in linear combinations that associate with energy density highlights. Mathematically, the problem is to find an orthonormal transformation,  $\mathbf{C}$ , such that

$$\mathbf{y}_i = \sum_j c_{i,j} \mathbf{u}_j \quad (5)$$

Where  $\mathbf{y}_i$  are the rotated basis set of vectors that associate with the  $\mathbf{u}_i$ . Also the requirement is imposed that the resulting  $\mathbf{y}_i$  are orthonormal, i.e., the inner product is such that  $(\mathbf{y}_i, \mathbf{y}_j) = \delta_{i,j}$ . This implies that

$$\mathbf{C}\mathbf{C}^T = \mathbf{C}^T\mathbf{C} = \mathbf{I} \quad (6)$$

In equation (5) both the  $\mathbf{y}_i$  vectors and the  $c_{i,j}$  coefficients are unknowns. When the  $\mathbf{A}$  matrix of equation (3) is  $(M \times N)$ , then SVD yields an  $(M \times M)$   $\mathbf{U}$  matrix whose columns are the vectors  $\mathbf{u}_i$ , and an  $(N \times N)$   $\mathbf{V}$  matrix whose columns are the vectors  $\mathbf{v}_i$ , and an  $(M \times N)$   $\mathbf{S}$  matrix whose diagonal elements are the singular values. The solution to the problem posed by equation (5) is to find the  $c_{i,j}$  coefficients in some optimum fashion. This was accomplished as follows: first, the means of the  $\mathbf{y}_i$  vectors are formulated in terms of the  $c_{i,j}$  coefficients. The means for the  $\mathbf{y}_i$  are

$$\begin{aligned} \langle m \rangle_i &= \frac{\sum_{r=1}^M m y_i^2(r)}{\sum_{r=1}^M y_i^2(r)} = \sum_{r=1}^M m y_i^2(r) \\ \langle m \rangle_i &= \sum_{r=1}^M m \sum_{s=1}^R c_{s,i} u_s(r) \sum_{t=1}^R c_{t,i} u_t(r) \\ \langle m \rangle_i &= \mathbf{c}_i^T \mathbf{M} \mathbf{c}_i \end{aligned} \quad (7)$$

It is fortuitous, as it turns out, that the means of the transformed vectors are of a quadratic form. This

provides a unique, optimum solution for the  $c_{i,j}$  coefficients. The extrema for the means of the  $\mathbf{y}_i$  vectors are achieved when the  $\mathbf{c}_i$  are the eigenvectors of the  $\mathbf{M}$  matrix of equation (7). In similar fashion, an orthonormal transformation matrix  $\mathbf{D}$  can be found for the  $\mathbf{v}_i$  vectors such that

$$\mathbf{x}_i = \sum_j d_{i,j} \mathbf{v}_j \quad (8)$$

$$\mathbf{D}\mathbf{D}^T = \mathbf{D}^T\mathbf{D} = \mathbf{I}$$

and

$$\langle n \rangle_i = \mathbf{d}_i^T \mathbf{N} \mathbf{d}_i \quad (9)$$

where the  $\mathbf{d}_i$  are the eigenvectors of the  $\mathbf{N}$  matrix.

When TSVD is applied only to the principal elements of the SVD of a matrix, the resulting series pertains to the prominent amplitude distribution of the original matrix.

The following summarizes the basis decomposition methods -- SVD and the new TSVD method:

**Matrix Decomposition by SVD Method:**

$$\mathbf{A} = \mathbf{U}\mathbf{S}\mathbf{V}^T$$

**Matrix Decomposition by TSVD Method:**

$$\mathbf{U} = \mathbf{X}\mathbf{C}^T$$

$$\mathbf{V} = \mathbf{Y}\mathbf{D}^T$$

Therefore,

$$\mathbf{A} = (\mathbf{X}\mathbf{C}^T)\mathbf{S}(\mathbf{Y}\mathbf{D}^T)^T$$

$$\mathbf{A} = \mathbf{X}(\mathbf{C}^T\mathbf{S}\mathbf{D}^T)\mathbf{Y}$$

#### IV. Example and Discussion

This example presents a simple illustration of the new TSVD method for decomposing a matrix and extracting the principle features. The signal of interest is a series of five equal amplitude sin bursts, each at a separate frequency, namely, 20, 15, 10.5, and 25 normalized frequency units. Figure 1 presents the time series for this signal. Figure 2 is the spectrogram for the signal. Notice the distinct, equal amplitude energy density highlights at the appropriate time-frequency locations in the time-frequency plane. An SVD was performed on the spectrogram matrix. Figures 3 and 4 present the first five (principal)  $\mathbf{u}_i$  and  $\mathbf{v}_i$  vectors respectively. The  $\mathbf{u}_i$  associate with time and the  $\mathbf{v}_i$  associate with frequency. Note that no one  $\mathbf{u}_i$  vector can be attributed to a particular sin burst. Likewise, no one  $\mathbf{v}_i$  vector can be attributed to a given frequency for a particular sin burst. Figure 5 and 6 are the rotated basis vectors  $\mathbf{y}_i$  and  $\mathbf{x}_i$  that associate with time and frequency respectively. It is easy to see from figures 5 and 6 that the new basis vector set, the  $\mathbf{y}_i$  and  $\mathbf{x}_i$  vectors, derived by forming linear combinations of the SVD basis

vectors  $u_i$  and  $v_i$ , associate directly with the respective time and frequency locations of the energy density for the sin burst depicted in figure 2.

Since the coefficients for finding these rotated basis sets are derived from the directions of the extrema of the respective means of the  $y_i$  and  $x_i$  vectors, they are optimum in that sense. Further more, the salient

optimum principal (principal from SVD and optimum from new method) features can be directly obtained from the  $y_i$  and  $x_i$  vectors viz. the location in time and frequency and the spread in time and frequency for each energy density that associates with a particular sin burst.

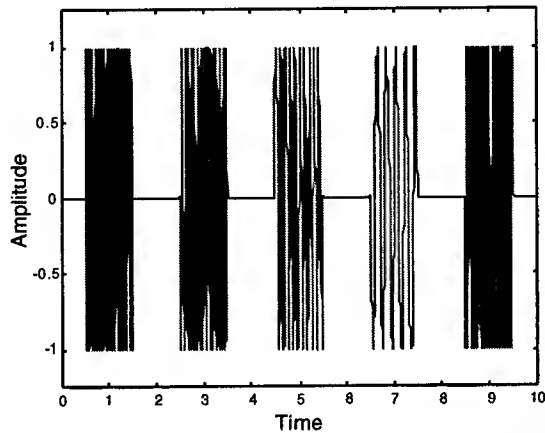


Figure 1 -- Time Series

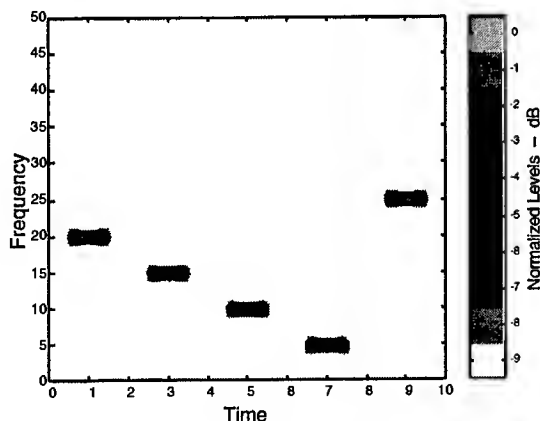


Figure 2 -- Spectrogram

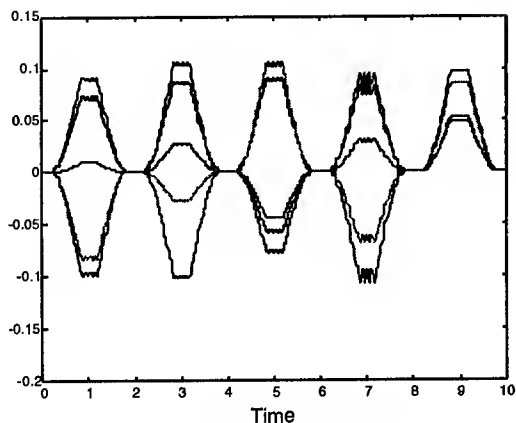


Figure 3 Singular Time Vectors from SVD

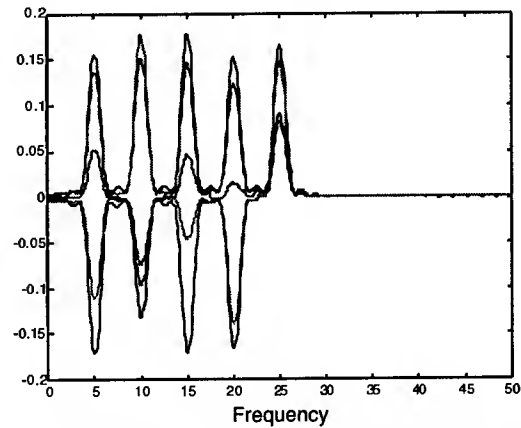


Figure 4 Singular Frequency Vectors from SVD

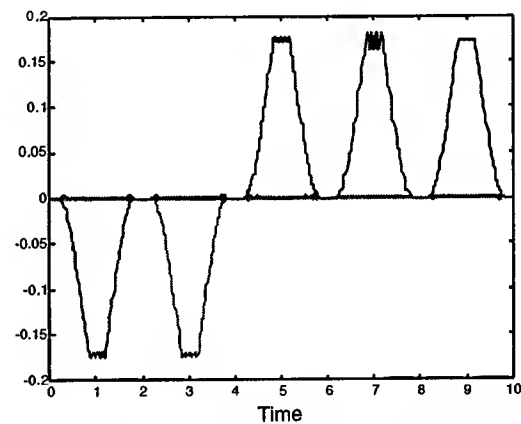


Figure 5 Rotated Time Vector Basis Set via New Matrix Decomposition

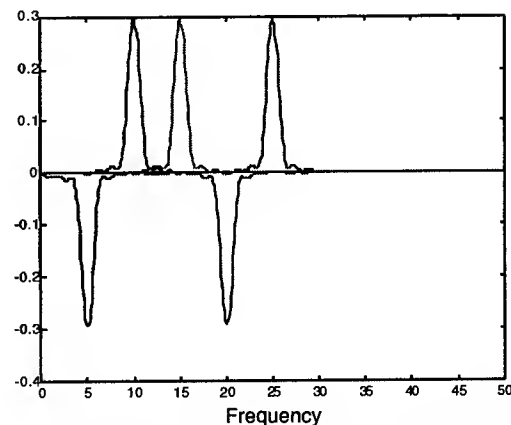


Figure 6 Rotated Frequency Vector Basis Set via New Matrix Decomposition

## References

- [1] E. Wigner, "On The Quantum Correction For Thermodynamic Equilibrium", *Phys. Rev.*, vol. 40, pp. 749-759, 1932.
- [2] E. Wigner, "Quantum-Mechanical Distribution Functions Revisited", *Perspectives in Quantum Theory*, W. Yourgrau and A. van der Merwe, Ed. Cambridge, MA; MIT Press, 1971.
- [3] L. Cohen, "Time-Frequency Distributions – A Review", *Proceedings of The IEEE*, vol. 77, No. 7, July 1989.
- [4] L. Cohen, *Time-Frequency Analysis*, Prentice-Hall, New York, 1995.
- [5] L. Cohen, "Generalized phase-space distribution functions," *J. Math. Phys.*, vol. 7 no. 5, pp. 781-786, 1966.
- [6] H. Choi and W. Williams, "Improved time-frequency representation of multicomponent signals using exponential kernels", *IEEE Trans. Acoust. Speech, Signal Processing*, vol. 37, pp 862-871, 1989.
- [7] Y. Zhao, L. Atlas, and R. Marks, II, "The Use of Cone-Shaped Kernels for Generalized Time-Frequency Representations of Nonstationary Signals", *IEEE Trans. Acoust, Speech Signal Processing*, vol. 38, pp. 1084-1091, July 1990.
- [8] P. Loughlin, J. Pitton and L. Atlas, "Construction of positive Time Frequency Distributions, " *IEEE Trans. Sig. Proc.*, vol. 42, no. 10, pp. 2697-2705, 1994.
- [9] Dale Groutage, "A Fast Algorithm for Computing Minimum Cross-Entropy Positive Time-Frequency Distributions", *IEEE Trans. Sig. Proc.*, vol. 45, No. 8, August 1997, pp 1954-1970.
- [10] P. Loughlin, *Proceedings of the IEEE*, special issue on "Time-Frequency Analysis: Biomedical, Acoustical and Industrial Applications", P. Loughlin (ed.), vol. 84 No. 9, 1996.
- [11] N. M. Marinovic and G. Eichmann, Feature Extraction and Pattern Classification in Space-Spatial Frequency Domain, Proc. SPIE, Vol. 579, Conf. on Intelligent Robots and Computer Vision, 15-20, September, 1985, Cambridge, MA
- [12] N. M. Marinovic and G. Eichmann, An Expansion of Wigner Distribution and its Applications, Proceedings of the IEEE ICASSP-85, Vol. 3, pages 1021-1024, 1985.
- [13] Dale Groutage and David Bennink, "Feature Sets For Non-Stationary Signals Derived From Moments Of The Singular Value Decomposition Of Cohen-Posch (Positive Time-Frequency) Distributions", to appear in may issue of *IEEE Trans. Acoust. Speech, Signal Processing*, May, 2000.
- [14] C. L. Lawson and R. J. Hanson, *Solving Least Squares Problems*, Prentice-Hall, Inc., Englewood Cliffs, N. J., 1974.

# MINIMUM ENTROPY TIME-FREQUENCY DISTRIBUTIONS

*Amro El-Jaroudi*

Department of Electrical Engineering  
University of Pittsburgh  
348 Benedum Hall  
Pittsburgh PA 15261 USA  
amro@ee.pitt.edu

## ABSTRACT

We present an approach to designing discrete time-frequency distributions that are extremely localized in the time-frequency plane. These distributions, which satisfy the marginals, are constructed recursively by transferring energy among the points in the time-frequency distribution (TFD) in a direction which decreases the entropy of the TFD. This transfer is such that the resulting TFD continues to satisfy the marginal constraints.

## 1. INTRODUCTION

Entropy in general and maximum entropy in particular has long been used in constructing power spectral densities [1]. A popular approach in the spectral analysis of stationary signals is to find the spectral density that maximizes an entropy-based criterion while satisfying autocorrelation matching constraints [1]. The result is the spectrum of an autoregressive filter whose coefficients are obtained by solving a linear set of equations. (It is important to note that the result depends greatly on the definition of entropy [2].) Proponents of maximum entropy have long argued that the resulting spectrum is "obtained while making the fewest assumptions about the signal" and "the flattest spectrum which satisfies the constraints" [3]. Maximum entropy also became synonymous with high resolution spectral estimation due to the nature of its results in the stationary case: peaky spectra based on all-pole filters. However, it is often difficult to reconcile the notion of "flattest possible spectrum" with that of a "high resolution spectrum" [3]. In time-frequency (TF) analysis, maximum entropy has been used to generate time-frequency distributions (TFD) which satisfy marginal constraints [5]. It has also been used as a criterion for deblurring an initial TFD and for estimating the evolutionary spectrum [4].

In time-frequency analysis, a goal is often to produce distributions that localize the energy density in the TF plane. Such goals seem in direct contradiction with the notion of "flattest possible distribution" often associated with maximum entropy. In fact, this goal is more in alignment with the concept of a minimum entropy distribution. In this paper, we present a method to construct minimum entropy discrete TFDs that satisfy marginal constraints. We demonstrate that this method is guaranteed to converge to a local minimum in a finite number of steps. The resulting TFDs are highly localized as measured by the decrease in entropy and the increase in the number of zero values in the TF plane.

## 2. APPROACH

The proposed algorithm starts with a TFD,  $P(n, \omega)$ , that satisfies the marginals. For example, for a unit energy signal  $x(n)$ , one can use the correlationless distribution

$$P(n, \omega) = |x(n)|^2 |X(\omega)|^2. \quad (1)$$

In fact, this distribution is the one which maximizes the entropy while satisfying the marginals. While it is a valid TFD, it produces little information beyond the marginals. The proposed algorithm can start with any positive distribution that satisfies the marginals (e.g., see [4,5]).

### 2.1 Basic principle

The basic principle behind the proposed method is to modify the initial TFD in a direction that decreases the entropy while not disturbing the time and frequency marginals. For this purpose, we select four points which form the corners of a rectangular grid in the T-F plane. Note that these points do not have to be adjacent, they merely

have to form the corners of a rectangle. An example is shown in Figure 1 where the four points are labeled  $p_{1,1}, p_{1,2}, p_{2,1}, p_{2,2}$ .

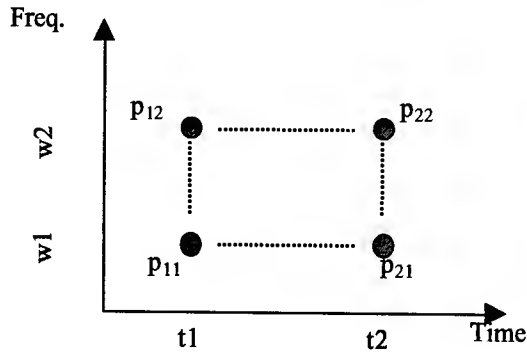


Figure 1. An example of a rectangular grid in the T-F plane

In the figure, the points  $p_{1,1}, p_{1,2}$  occur at time  $t1$  while the points  $p_{2,1}, p_{2,2}$  occur at time  $t2$ . ( $t1$  and  $t2$  need not be consecutive in value). Also  $p_{1,1}, p_{2,1}$  occur at frequency  $w1$  while  $p_{1,2}, p_{2,2}$  occur at frequency  $w2$  ( $w1$  and  $w2$  need not be consecutive either).

The proposed method is to adjust the TFD by subtracting a value  $\Delta$  from  $p_{1,1}, p_{2,2}$  and adding the same value to  $p_{1,2}, p_{2,1}$ , thereby creating the following new set of values  $(p_{1,1} - \Delta), (p_{1,2} + \Delta), (p_{2,1} + \Delta), (p_{2,2} - \Delta)$ . It is important to note that along  $t1$ , for example,  $\Delta$  is subtracted from  $p_{1,1}$  and added to  $p_{1,2}$ . Consequently, the sum of all the values along  $t1$  (the time marginal at  $t1$ ) does not change. Therefore, if the initial distribution satisfies the marginal at  $t1$ , it will continue to do so. It is easy to show that the same is true for  $t2$ , as well as for  $w1$  and  $w2$ . Since no other values in the TFD are disturbed, the modified TFD will have the same marginals as the initial one.

The remaining obstacle is to choose the optimal value of  $\Delta$  which will minimize the entropy of the resulting TFD. Since we limit ourselves to nonnegative TFDs, the value of  $\Delta$  is limited from above by the minimum of  $p_{1,1}, p_{2,2}$  and from below by the negative of the minimum of  $p_{1,2}, p_{2,1}$ . This guarantees that the new values  $(p_{1,1} - \Delta), (p_{1,2} + \Delta), (p_{2,1} + \Delta), (p_{2,2} - \Delta)$  are

nonnegative. Below we find the optimal value of  $\Delta$  over this range.

## 2.2 Optimal Choice for $\Delta$

Let the entropy of a TFD  $P(n, \omega)$  be given by

$$E = - \sum_n \int_{\omega} P(n, \omega) \ln P(n, \omega)$$

We can then define the change in entropy which occurs due to the modification of the four TFD points described above as

$$Loss = E_{before} - E_{after} \quad (2)$$

We would then choose  $\Delta$  to maximize the entropy loss. The *Loss* can be written in terms of only the four points which are modified since the contribution of the other points to the entropy is the same before and after the modification. In other words, (2) can be rewritten as

$$\begin{aligned} Loss = & -p_{1,1} \ln p_{1,1} - p_{1,2} \ln p_{1,2} \\ & -p_{2,1} \ln p_{2,1} - p_{2,2} \ln p_{2,2} \\ & + (p_{1,1} - \Delta) \ln (p_{1,1} - \Delta) \\ & + (p_{1,2} + \Delta) \ln (p_{1,2} + \Delta) \\ & + (p_{2,1} + \Delta) \ln (p_{2,1} + \Delta) \\ & + (p_{2,2} - \Delta) \ln (p_{2,2} - \Delta) \end{aligned} \quad (3)$$

All that remains is to maximize the above equation with respect to  $\Delta$ . (Note that the first four terms do not depend on  $\Delta$ . Consequently, they can be ignored during the maximization). Taking the derivative of the *Loss* with respect to  $\Delta$  and setting it to zero yields:

$$\begin{aligned} 0 = & -\ln(p_{1,1} - \Delta) + \ln(p_{1,2} + \Delta) \\ & + \ln(p_{2,1} + \Delta) - \ln(p_{2,2} - \Delta) \end{aligned} \quad (4)$$

The above equation is easily solved for  $\Delta$  to yield

$$\Delta = \frac{p_{1,1}p_{2,2} - p_{1,2}p_{2,1}}{p_{1,1} + p_{1,2} + p_{2,1} + p_{2,2}} \quad (5)$$

The second derivative of the loss function with respect to  $\Delta$  is given by

$$\begin{aligned} \frac{\partial^2 Loss}{\partial \Delta^2} = & \frac{1}{p_{1,1} - \Delta} + \frac{1}{p_{1,2} + \Delta} \\ & + \frac{1}{p_{2,1} + \Delta} + \frac{1}{p_{2,2} - \Delta} \end{aligned} \quad (6)$$



This quantity is strictly positive for the range of  $\Delta$  over which the optimization is performed. Unfortunately this implies that the critical value of  $\Delta$  in (5) minimizes the *Loss* in (3) instead of maximizing it. This also means that the entropy loss function in (3) is convex in  $\Delta$  and, to maximize it, one has to choose one of the extreme values of  $\Delta$ : the minimum of  $p_{1,1}, p_{2,2}$  or the negative of the minimum of  $p_{1,2}, p_{2,1}$ . The choice is made by evaluating the loss function in (3) for the two possible values of  $\Delta$  and choosing the one which causes the larger loss.

It is interesting to note that after the modification, one of the four points will become identically zero. Consequently, the approach reduces the number of nonzero values in the time-frequency plane, i.e., it concentrates the energy distribution into fewer time-frequency points, thereby achieving higher localization.

It is also interesting to note that, if our goal were to maximize the entropy of the resulting TFD, we could then use the  $\Delta$  in (5) to update the values in the TFD. Moreover, if we start with a maximum entropy TFD (using (1)), it is easy to show that the  $\Delta$  in (5) is identically zero. This is intuitively appealing since we cannot expect to increase the entropy of a TFD that is already maximum entropy.

### 3. ALGORITHM

In this section we describe the algorithm that implements the minimum entropy approach described above and comment on its performance and limitations.

#### 3.1 Algorithm Steps

1. Choose a pair of points  $p_{1,1}, p_{2,2}$  in the T-F plane. This will define a rectangle as long as the two points do not occur at the same time or at the same frequency. Consequently the choice of  $p_{1,1}, p_{2,2}$  defines the set  $p_{1,1}, p_{1,2}, p_{2,1}, p_{2,2}$ .
2. Evaluate the loss function in (3) for the two possible values of  $\Delta$ :  $\min\{p_{1,1}, p_{2,2}\}$  or  $-\min\{p_{1,2}, p_{2,1}\}$ . Choose the value that results in a larger loss.
3. Update the set  $p_{1,1}, p_{1,2}, p_{2,1}, p_{2,2}$  using the  $\Delta$  from step 2 to create the set of points  $(p_{1,1} - \Delta), (p_{1,2} + \Delta), (p_{2,1} + \Delta), (p_{2,2} - \Delta)$ .

4. If there is no remaining pair of points that will reduce the entropy, stop. Otherwise return to step 1.

### 3.2 Comments

The algorithm above is computationally simple and requires very few operations at each step. Unfortunately, the number of steps, i.e., the number of possible pair of points in the T-F plane, is very large (on the order of  $N^4$ , assuming  $N$  points in time and  $N$  points in frequency). This computational cost makes the algorithm very cumbersome. On the other hand, the algorithm is easily parallelized.

An important factor that affects the resulting TFD is the order by which the pairs of points are chosen in step 1 of the algorithm. Each ordering produces a different result. Consequently, care should be taken in choosing the sequence of pairs. One logical choice is the pair which achieves the greatest reduction in entropy. However this choice involves a search through all the possible choices which increases the computational burden of the algorithm.

## 4. RESULTS

To test the algorithm, we use a chirp of length  $N=32$  given by the following equation

$$x(n) = ce^{j\pi n^2 / (2N)}$$

where  $c$  is a normalization constant to make the signal unit-energy. Figure 2 shows the real part of the signal. Figure 3 shows the initial TFD (calculated using (1)) which satisfies the marginals and has maximum entropy. Figure 4 shows the minimum entropy TFD calculated using the proposed algorithm. The final TFD has entropy 3.8 compared to the initial TFD's entropy of 6.4. Moreover, the final TFD concentrated the energy in 209 nonzero time-frequency points compared to 1024 for the initial TFD. As expected, the final TFD also satisfies the marginals.

It is important to remember that the algorithm converges to a local minimum of the entropy depending on the sequence of points chosen in step 1 above. No effort was made in this example to optimize the sequence.

While the appearance of the final TFD may not be pleasing to some readers, it is nonetheless a valid TFD which satisfies the marginals and is highly

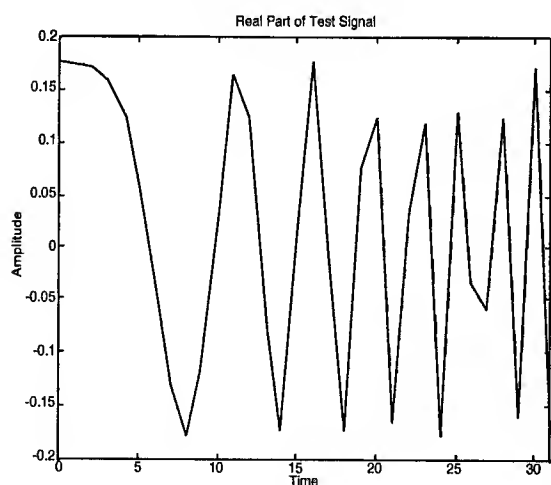
localized in time-frequency. Also the final appearance depends greatly on the initial TFD. If, for example, we start with a TFD concentrated along the instantaneous frequency (IF), the final result will be a more localized TFD also concentrated along the IF. One may also consider additional constraints to impose on the resulting TFD on top of the marginal constraints. However these additional constraints may require the modification of the algorithm presented in Section 3.

## 5. CONCLUSIONS

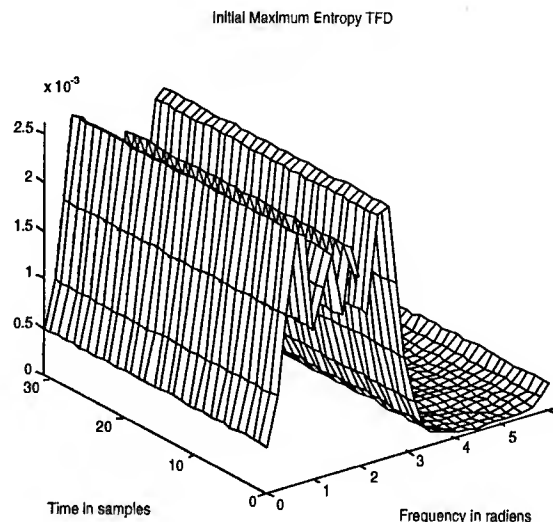
We presented an algorithm to calculate a minimum-entropy TFD that satisfies the marginals. The resulting TFDs are highly localized in the TF plane.

## 6. REFERENCES

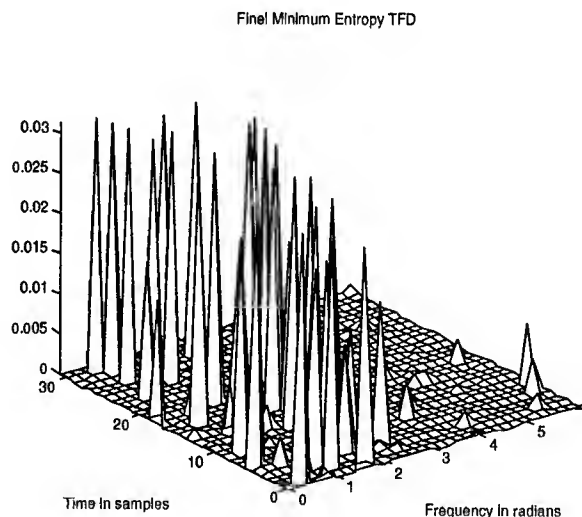
- [1] S. Kay, *Modern Spectral Estimation*, Prentice Hall, Englewood Cliffs, N.J., 1988.
- [2] C. Nadeu, "Maximum Flatness Spectral Modeling", *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 38, no. 11, pp., 2006-2008, November 1990.
- [3] J. Makhoul, "Maximum Confusion Spectral Analysis", *Proc. 3<sup>rd</sup> ASSP Workshop on Spectral Estimation and Modeling*, Boston, MA, November 1986, pp. 6-9.
- [4] J. Pitton, L. Atlas and P. Loughlin, "Applications of positive time-frequency distributions to speech processing," *IEEE Transactions on Speech and Audio Processing*, vol. 2, no. 4, pp. 554-566, 1994.
- [5] P. Loughlin, L. Atlas, and J. Pitton, "Construction of Positive Time-Frequency Distributions", *IEEE Transactions on Signal Processing*, vol. 42, no. 10, pp. 2697-2705, 1994.



**Figure 2** Real part of test signal



**Figure 3** Initial maximum-entropy TFD



**Figure 4** Final minimum-entropy TFD

# UNCERTAINTY IN THE TIME-FREQUENCY PLANE

Paulo M. Oliveira (\*) and Victor Barroso (\*\*)

(\*) Escola Naval - DAEL, Base Naval de Lisboa, Alfeite, 2800 Almada, Portugal

(\*\*) Instituto Superior Técnico ISR/DEEC, Torre Norte, Piso 7 Av. Rovisco Pais, 1049-001, Lisboa, Portugal

## ABSTRACT

Is there a limit to the maximum resolution one can achieve when representing the signal's energy in the Time-Frequency plane? Some authors sustain that such a limit exists, and ignoring it is the cause of the known difficulties with some joint Time-Frequency distributions; others maintain that there is no such limit.

In this article, we propose to analyze the merits and demerits of the several existing approaches, and suggest further arguments one might wish to consider. This will take us to the conclusion that, both from a tool-specific and from a general information-theoretic point of view, there is, indeed, a lower limit on the achievable resolution, even though the expression for that limit can not be given by the traditional Heisenberg-Gabor relations.

## 1. INTRODUCTION

The implications of the so called "Principle of Uncertainty" in the signal processing field have first been recognized by Gabor [1], who introduced a time-frequency version of Heisenberg's inequality:

$$\sigma_t \cdot \sigma_f \geq \frac{1}{4\pi}, \quad (1)$$

where  $\sigma_t$  and  $\sigma_f$  are the time and frequency standard deviations, respectively. It is often assumed that this relation implies the existence of a maximum possible resolution within the time-frequency plane. It does not, as will be seen, and has often been pointed out [2]. But is there such a limit, even if (1) fails to address it? Answering this question is the goal of this article. We will approach the objective in steps. We will start by looking at the existing uncertainty relations and other traditional approaches. After having established their shortcomings, we will propose a different and more general approach, one which will lead us to the desired answer.

## 2. THE MATHEMATICAL RELATION

As a mathematical relation, (1) is a very simple statement, concerning the standard deviations of a function and its

Fourier Transform. As a matter of fact, it is not the best one, since stronger statements can be made [2]:

$$\sigma_t \cdot \sigma_f \geq \frac{1}{2\pi} \sqrt{\frac{1}{4} + Cov^2}, \quad (2)$$

where  $Cov$  is the covariance of the signal, defined in [2]. Since  $Cov^2$  is necessarily a non-negative quantity, (2) is stronger than (1).

Both relations (1) and (2) must, however, be taken carefully, since their misinterpretation and misuse is responsible for many false common notions. • Firstly, we note that they fail to express what one usually feels to be the uncertainty principle. Because they use standard deviations as measures of spread, they do not prohibit the existence of functions arbitrarily narrow in both time and frequency domains. In fact, it is always possible to devise a function arbitrarily narrow (in the sense that its energy can be made to be arbitrarily concentrated) and whose standard deviation is greater than any given value [3]. • Another inadequacy of the use of standard deviations (or any other measure of global width) appears when the signal  $s(t)$  (or its spectrum  $S(f)$ ) is not unimodal [3]. If, for example, we think of a finite segment of a two-tone signal, we will easily conclude that the standard deviation of  $S(f)$  is, for any reasonably long observation time, almost independent of the length of the observation period. Instead, it depends strongly on the individual frequencies of the two tones. Once again, (1) and (2) fail to express what we intuitively feel the uncertainty principle to be in this particular case: an inverse relation between the duration of the observed segment, and the (local) width of the main lobes of its power spectrum. • As a last comment, we will note that these relations fail to express reciprocity between both domains (an increase in  $\sigma_t$  does not necessarily imply a corresponding decrease in  $\sigma_f$ ), a fact which has often been pointed out as one of the limitations of their quantum mechanical counterparts (e.g. [4]).

Attempts have been made to obtain alternative measures of width in time and frequency, that might better express the concept of narrowness. One of the most successful approaches was the one of Slepian-Landau-Pollack who, in a series of papers (e.g. [5]), showed that it is not possible to design a function with arbitrary simultaneous energy concentration in arbitrarily small regions of time and frequency. The particular limits on energy concentration depend on the desired time-bandwidth product [6], but the important point is that they exist. Let us just note that this broadness

This work was supported by the Ministério da Defesa Nacional and Fundação das Universidades Portuguesas, subprogram "Os Oceanos e as suas Margens".

measure also fails to express what one feels to be the uncertainty principle in the case of non-unimodal signals (or with non-unimodal spectra).

We will use a different definition of broadness, based on the Hessian of the power spectrum. Define the broadness ( $B_P$ ) of a power spectrum  $P(f) = |S(f)|^2$  as:

$$B_P = 1/\sqrt{Max\left(-\frac{\partial^2}{\partial f^2}P(f)\right)}. \quad (3)$$

When applied to the frequency domain, this type of measure relates directly to the idea of frequency resolution. As can be expected and is easily shown, it also obeys an "uncertainty relation":

$$B_P \cdot \sigma_R \geq \frac{1}{2\pi}, \quad (4)$$

where  $\sigma_R^2 = \int_{-\infty}^{\infty} \tau^2 |R(\tau)| d\tau$ , and  $R(\tau) = \int_{-\infty}^{\infty} s(t)s^*(t-\tau) dt$ . It is, however, a relation between the broadness of the fine structure in one domain and a global width (standard deviation) in the other. It thus conveys the general feeling of what the "uncertainty principle" is, better than the relations between the global widths of both domains, such as (1). Namely, it can be used with non-unimodal spectra such as the two-tone signal previously considered. There are other interesting characteristics in (3). One of them is that the fine structure of the spectrum is seen to be related to the overall width of the autocorrelation function, and not to the time duration of the signal, a point which we will soon explore. Another interesting fact is that it possesses a counting property (similar to the ones investigated in [7] and [8] for the Rényi entropy), an useful attribute that will not be explored here. Equivalent relations can, of course, be obtained for the complementary domain (fine structure in time *vs* spectral overall width). We may now question if there is any uncertainty relation between the fine structure in one domain and the fine structure in the other. There is not, as can easily be shown.

Other definitions of broadness can be devised. An interesting one (which we will not pursue here) is:

$$B_R = \frac{1}{\sqrt{-\frac{\partial^2}{\partial f^2} \left( \int_{-\infty}^{\infty} P(f - \frac{v}{2}) P(f + \frac{v}{2}) df \right) \Big|_{v=0}}}, \quad (5)$$

since, with this definition,  $B_R \cdot \sigma_R = 1/2\pi$ . This measure of broadness thus creates reciprocity between the widths in the dual domains, something which (1) or (2) can not achieve, as already mentioned.

### 3. THE PHYSICAL PRINCIPLE

An attempt to grab the physical principle behind these different mathematical manifestations was made by Robertson [9] with a simple but apparently far-reaching statement: there will be uncertainty between two variables (observables) whenever their operators don't commute. Mathematically, if  $\mathcal{A}$  and  $\mathcal{B}$  are the operators associated with the variables, Robertson's inequality states that:

$$\sigma_t^2 \cdot \sigma_f^2 \geq \frac{1}{4} |\langle \mathcal{AB} - \mathcal{BA} \rangle|^2,$$

where  $\langle \cdot \rangle$  stands for average. This apparent generalization (its role as a generalization has been severely criticized - see, for example [10]) of (1) does not, however, shed much light on the underlying physical principle, since it relies too heavily on the mathematical formalism of operators, a technique of which Nature knows nothing. Uncertainty relations have been obtained in physical areas where the mathematical formalism of operators can not even be applied (e.g. [11]). One must think of the uncertainty principle as a consequence of the eigenfunctions of the chosen operators (and, thus, of the definition of the "pure" representatives of the physical quantities). In the time-frequency case, there will be uncertainty if the eigenfunctions corresponding to the concept of "frequency" do not have time localization properties. The existence of uncertainty between the dual domains is, thus, not an exclusive of the Fourier Transform. As an example of that, let us choose constant amplitude chirps as the "pure" representatives of the frequency concept (assuming that there is any sense whatsoever in doing this). Formally, this would correspond to define "frequency" as the eigenvalues of the operator  $\Upsilon$ , where:

$$\Upsilon = - \left( 2\alpha_0 t - \frac{1}{j2\pi} \frac{\partial}{\partial t} \right),$$

the eigenfunctions being  $e^{j(\alpha_0 t^2 + 2\pi f t)}$ . The "frequency" eigenfunctions have, thus, no time localization capabilities. Is there an associated uncertainty principle between time and "frequency"? Yes. As is easily shown, the mathematical relation is again (1). Instead of linear chirps, we could do the same for quadratic chirps, or constant amplitude functions of any polynomial phase law (and, thus, extendable to more general signals via the Weierstrass theorem), and the conclusions will be the same. This means, namely, that uncertainty relations will exist between time and "frequency" for the fractional Fourier Transform with any angle of rotation in the time-frequency plane.

Avoiding uncertainty relations implies an acceptable redefinition of the involved concepts. If, for instance, one is willing to redefine *frequency*, accepting the concept of *local frequency* as a suitable physical quantity, and thus using time localized waves as representatives of the "pure" concept, then the uncertainty relations are easily made to collapse. Let us use, as an example, cisoids with a Gaussian envelope:

$$\varphi(t, f) = e^{-\frac{t^2}{4\sigma_t^2}} e^{j2\pi f t}.$$

These "pure" functions are the eigenfunctions of the non-Hermitian (and non-linear) operator

$$\mathcal{F} = \frac{1}{2\pi} Im \left\{ \frac{\partial}{\partial t} \ln[\cdot] \right\},$$

whose (continuous) eigenvalues are the values of  $f$ , thus carrying with them the concept of local frequency. Let us now consider a signal whose (local) frequency spectrum is

$$S(f) = e^{-\frac{f^2}{4\sigma_f^2}}.$$

That is:

$$s(t) = \int_{-\infty}^{\infty} S(f) \varphi(t, f) df =$$

$$= \int_{-\infty}^{\infty} e^{-\frac{f^2}{4\sigma_f^2}} e^{-\frac{t^2}{4\sigma_e^2}} e^{j2\pi ft} dt.$$

Denoting the variance of  $s(t)$  by  $\sigma_t^2$ , we have that [12]:

$$\sigma_t^2 = \frac{1}{\left(\frac{1}{\sigma_e^2} + 16\pi^2\sigma_f^2\right)}.$$

For any given  $\sigma_f^2$ , we can now arbitrarily diminish the time width by simple decreasing  $\sigma_e^2$ . The uncertainty relations have, in fact, collapsed. In the limit of no time localization capabilities on the frequency concept ( $\sigma_e^2 \rightarrow \infty$ ), these relations reduce back to (1), as they had to.

#### 4. THE TIME-FREQUENCY PLANE

Inequalities like (1) and (2) impose restrictions on the simultaneous behavior of a time function and its Fourier transform. But imposing conditions on the time and frequency marginals doesn't really tell us much about what is or is not achievable within the time frequency plane [13]. As an example, consider the time-frequency representation  $\rho(t, f) = \delta(f - kt)$ . Obtaining such a time-frequency representation implies having infinite local time-frequency resolution. And yet, for any given  $\sigma_t^2$ , one can force the frequency marginal to be arbitrarily wide by simply increasing the chirp rate. Hence, (1) does not, in fact, constitute a limit to the achievable time-frequency resolution *within* the time-frequency plane. This is also true for all uncertainty relations discussed so far.

To determine the limits of joint time-frequency descriptions, we need to abandon the marginals, with their global time and Fourier descriptions, and move into the plane. Some attempts have been made, mainly using the Wigner-Ville Distribution (or, more generally, the Cohen class of distributions) as a joint energy description of the signal. A very good summary of these can be found in [6]. However, careful reasoning (the analysis of each one of these proposals can not be done here, for space reasons) will show that none of them addresses properly the issue of time-frequency concentration. A more promising approach is the extension of the Slepian-Pollack-Landau energy concentration measure to ellipsoidal regions with axis parallel to the time and frequency axis [6]. Unfortunately, the results are specific to the bilinear class and, furthermore, this particular shape of the concentration region fails to answer the question in cases with spectral dynamics.

Other approaches have been made, considering conditional (and, thus, local) moments (e.g. [2], [14]). They concluded that (1) does not constrain the product of the local moments. Not being limited by (1) is, however, different from not being limited at all. That approach is, thus, inconclusive.

#### 5. SPECTRAL COMPLEXITY

To properly address the issue of joint time-frequency resolution, let us consider it from the more general point of view of information gathering. Let us first define the concept of *spectral complexity* of a stochastic signal ( $C_P$ ), as being

the amount of information one must collect to properly estimate its power spectrum (for the moment, let us assume that the signal possesses no spectral dynamics). Define

$$C_P = \sqrt{\text{Max} \left( -\frac{\partial^2}{\partial f^2} E\{P(f)\} \right)}, \quad (6)$$

where  $E\{P(f)\}$  is the power spectrum of the signal. Note that this measure of complexity is very similar to the definition of Fisher information, and bears a very close link with the notion of "narrowness" of the power spectrum.

To avoid the hassle of having to invoke or discard ergodicity, and so contribute to obscure the main issue, let us assume that we have a second degree "ensemble view". That is: by observing the signal at time  $t_1$ , we immediately apprehend the value of  $E\{s(t_1)\}$ ; by also observing the signal at time  $t_2$ , we now not only apprehend the value of  $E\{s(t_2)\}$ , but also  $R(t_1, t_2) = E\{s(t_1)s^*(t_2)\}$ . Since we are, under this assumption, directly observing expected values (and not mere realizations of the process), we may, in what follows, safely ignore the practical aspects of real estimators, such as bias and variance.

• **Stationary signals.** From this idealized point of view, let us now consider the estimation of the power spectrum of a stationary signal. To perform the estimate, one must extract information out of the signal. But how much observation time do we need? How much information must we collect? At first, increased observation time will provide better estimates, in a process converging to the true power spectrum. But, after convergence (assuming it ever happens), will further increases in observation time provide more information about the power spectrum? It clearly doesn't. Once this convergence process is completed, no further observation time is needed; no further information is required. We will have reached the signal's complexity  $C_P$ . If the observation time is less than the time to reach complexity ( $\tau_R$ ), we will be missing part of the information needed for a correct estimate; if the observation time available is greater than  $\tau_R$ , the last part of the signal will be informationless. But how do we determine  $\tau_R$ ? The amount of information needed to estimate the power spectrum is clearly the same amount of information needed to estimate its inverse Fourier Transform, the autocorrelation function. Hence, we only need to observe the signal for the amount of time needed to determine all (relevant) lags of its autocorrelation function. The time to reach complexity is thus the time support of the autocorrelation function. This is a very gratifying conclusion, since, from (3) and (4), the spectral complexity (6) and the time support of the autocorrelation function  $R(\tau)$  are, in fact, directly related to each other through yet another "uncertainty relation":

$$\frac{1}{C_P} \cdot \sigma_R \geq \frac{1}{2\pi}. \quad (7)$$

Denoting by  $D(t)$  the density of information contained in the signal, and by  $I_P(t)$  the amount of collected information, the collection procedure can be summarized as follows:

$$I_P(t) = \int_{t_0}^t D(\tau) d\tau \leq \int_{t_0}^{t_0 + \tau_R} D(\xi) d\xi = C_P. \quad (8)$$

From (6), the spectral complexity of a sinusoid is infinite. In fact, the power spectrum of a pure sinusoid will always become narrower with increasing observation time, without ever stabilizing. Complexity will never be reached for finite observation times. Coherently, the autocorrelation function is known to have infinite time support. Even though an impulsive spectrum may seem simple, we must note that the necessarily perfect localization needed for a proper estimate does require collecting an infinite amount of information. This is thus the high end of spectral complexity, where all observation time becomes useful and brings additional information.

In the low end, we have white noise. An instantaneous ensemble observation fully characterizes its autocorrelation function and, hence, its very low complexity (zero, in fact) power spectrum. Further observation of the noise will not contribute with any new information concerning its power spectrum.

In the general case, we have that signals with narrow-band components (and, thus, of high spectral complexity) require a high collection time  $\tau_R$ . Signals without narrow-band components (hence, of low spectral complexity) have small collection times. In any case, observing the ensemble for periods longer than  $\tau_R$  is not useful, since no additional information about the power spectrum will be obtained.

• **Non-stationary signals.** Assume that we want to estimate the power spectrum of a non-stationary signal at time  $t_1$ . This spectrum will have a given amount of spectral complexity ( $C_P$ ), and to properly estimate it, we need to collect this very same amount of information about the spectrum (or the autocorrelation function) at time  $t_1$ . But to represent time  $t_1$ , all we have is  $s(t_1)$  itself, and no finite amount of spectral information can be extracted from an instantaneous value of the signal. Information collected at times other than  $t_1$  will only be useful if and only if it is correlated with the spectral information at time  $t_1$ . In the previous case of stationary signals, the spectral information at any time was totally correlated with the spectral information at any other time. In the non-stationary case, however, we must weight the collected information with the non-unitary correlation factor (we will denote it by *utility factor* -  $u(t)$ ) that determines how useful is the collected information for estimates at time  $t_1$ . We now have to distinguish between useful past and future, and non-useful past and future. The collection procedure (8) becomes, using superscripts to denote the particular time for which the spectrum estimate is intended:

$$I_P^{t_1}(t) = \int_{t_0}^t D(\tau) u(\tau - t_1) d\tau, \quad \text{with}$$

$$I_P^{t_1}(t) \leq C_P^{t_1} = \int_{-\infty}^{\infty} D(\xi) u(\xi - t_1) d\xi. \quad (9)$$

This utility factor is thus just formalizing the fact that observing a non-stationary signal *now* does not tell us much concerning the spectrum of the signal a fortnight ago. The exception lies, of course, in the case of signals with deterministic and known frequency dynamics, since in these cases the information collected at any time can always be made useful, by taking the dynamics into proper account. Knowledge of the frequency dynamics thus makes the util-

ity factor constant and unitary, bringing the case of non-stationary signals to the very same situation one encounters with stationary signals.

As an example, consider the estimation of the power spectrum of a constant amplitude linearly chirping signal with uniformly distributed random phase:

$$s(t) = e^{j(\alpha t^2 + \theta)}.$$

Its autocorrelation function is easily seen to be:

$$R(t, t - \tau) = R(t, \tau) = e^{-j(\alpha \tau^2 - 2\alpha \tau t)}.$$

To determine the utility factor, we can now determine how correlated are the autocorrelation functions at different times  $R_R(t_2, t_1)$ ,  $t_2 > t_1$ . Due to the infinite energy of these autocorrelation functions, in the computation of their correlation factor we will limit the integration region to an arbitrarily large region centered at the zero lag. That is:

$$u(t_2 - t_1) = R_R(t_2, t_1, l) =$$

$$= \frac{1}{2l} \int_{-l}^l R(t_2, \tau) R^*(t_1, \tau) d\tau.$$

This means that, in our case,

$$u(t_2 - t_1) = \frac{\sin(2\alpha l(t_2 - t_1))}{2\alpha l(t_2 - t_1)}. \quad (10)$$

The inclusion of  $u(t)$  in (9) (in this case, a sinc function) will limit the amount of collectable information relative to time  $t_1$  and, thus, will upper bound the achievable spectral complexity and, hence, the achievable frequency resolution.

This is thus the answer we have been trying to obtain. There are limits to the achievable frequency resolution within the Time-Frequency plane, due to the fact that the period of time during which one can collect information concerning the spectrum at a given time is diminishing as the spectral dynamics increases. For increasing dynamics ( $\alpha$ , in our example) the useful neighborhood (and, hence, the amount of useful information) will continuously decrease, and so will the achievable spectral complexity. This namely means that the faster a chirp moves, the broader it becomes in the t-f plane. This predicted broadening of the power spectrum with the increase of the chirping rate is, in fact, observed in many bilinear time-frequency distributions (e.g. Rihaczek, Margenau-Hill, Page, etc.). To illustrate it, we computed the Margenau-Hill distribution of a cubic chirp. The results can be seen in Figure 1.

Let us now try to determine, in the case of our chirp, what is the best observation time. From (10), we see that the best strategy is to limit the observation time to the main lobe of the sinc function. That is, observe the signal between  $t_1 - \tau$  and  $t_1 + \tau$ , where  $\tau = \pi/2\alpha l$ . But this implies that  $\tau$  is the maximum lag of the observed autocorrelation function. That is, in this best case,  $l = \tau$ . From where we conclude that the best observation time for a linear chirp is

$$\tau = \sqrt{\frac{\pi}{2\alpha}} = \sqrt{\frac{1}{2 \frac{\partial}{\partial t} [f_i(t)]}}, \quad (11)$$

$f_i(t)$  being the chirp instantaneous frequency (in this particular case, we may safely identify the concept of instantaneous frequency with the derivative of the phase function).

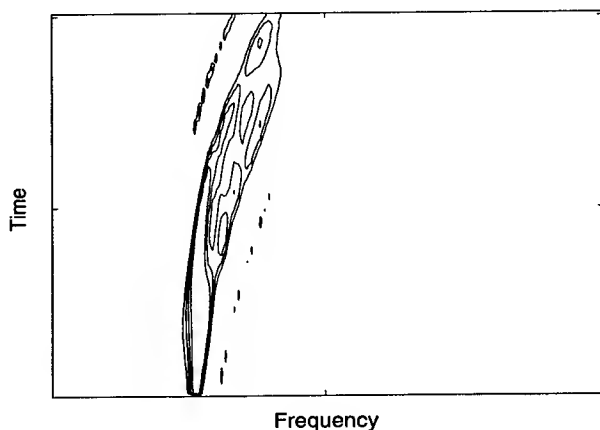


Figure 1: Cubic Chirp. Margenau-Hill distribution.

With hindsight, it is now interesting to observe that • (11) was already known to be the optimum observation time for short-time Fourier analysis of a chirp [15]; • (11) is the effective time support of the optimum data independent smoothing window to use with the Wigner-Ville distribution [16]; • it is also basically the same quantity defined by Rihaczek as the signal's "relaxation time" [17]. These separate results can now easily be understood as particular manifestations of (11).

A last comment must be made, concerning the use of models. Assuming a model for the frequency dynamics, such as the linear model implicit in the Wigner-Ville Distribution (or higher order models in the Polynomial Wigner-Ville Distribution), is an attempt to increase the size of what we called the "useful neighborhood", by projecting all collected spectral information to the time of interest. a) If, by inspiration or mere chance, the assumed model is, in fact, the correct one, we will overcome the limits imposed by the nonstationarity, and fall within the traditional, stationary uncertainty relations, as previously discussed; b) If, on the other hand, the model is incorrect, we must be prepared to pay for the wrong assumption. We will have apparently improved our frequency resolution, but we must pay for it with bias and artifacts. Another, more subtle, type of assumption, is made whenever we arbitrarily decide that the "true" distribution is the one maximizing some chosen criteria. It may or may not be a sensible, supported assumption. It is an assumption, nonetheless. It may buy us a better frequency resolution, if verified by the signal under analysis. In all other cases, one will pay for the apparently increased time-frequency resolution in bias and artifacts. This is, besides, exactly the type of trade-off one finds in all types of parametric spectrum estimation.

## 6. CONCLUSION

In this article, we addressed the issue of determining if there are lower bounds to the achievable time-frequency resolution within the Time-Frequency plane. After the analysis of existing approaches, we proposed an alternative one, based on the informational aspects of the estimation, in an attempt to achieve results independent of the specific

tool used to estimate the joint power spectrum. We concluded that there are, indeed, limits to the achievable time-frequency resolution. These limits are a direct consequence of the spectral dynamics of the signal. Increasing spectral dynamics imply decreasing time-frequency resolution capabilities. In the process of obtaining these limits, we also determined the optimum observation time (which also depends on the spectral dynamics of the signal), concluding that its optimality is tool independent. This conclusion allows an unified view of previously reported particular cases where this observation time was determined to be optimum.

## 7. REFERENCES

- [1] D. Gabor, "Theory of Communication," *Journal of the IEE*, Vol. 93, pp.429-457, 1946.
- [2] L. Cohen, *Time Frequency Analysis*. Englewood-Cliffs, NJ: Prentice-Hall, 1995.
- [3] J. Hilgevoord and J.B.M. Uffink, in *Microphysical Reality and Quantum Formalism*, Ed. A. van der Merwe et al., Kluwer Academic Publishers, 1988.
- [4] P. Kirschenmann, "Reciprocity in the uncertainty relations", *Philosophy of Science*, 40, 52-58, 1973
- [5] D. Slepian and H. Pollak, "Prolate spheroidal wave functions, Fourier analysis and Uncertainty - I, *Bell Syst. Tech. J.*, 40, pp. 43-64.
- [6] P. Flandrin, *Time-Frequency/Time-Scale Analysis*, Academic Press, 1999.
- [7] W. J. Williams, M. L. Brown, A. O. Hero III, "Uncertainty, information and time-frequency distributions," *Proc. Adv. Sig. Proc. Alg. Arq. and Implem. II*, pp. 144-156, San Diego, July 1991.
- [8] P. Flandrin, R. Baraniuk, O. Michel, "Time-Frequency Complexity and Information," *Proc. ICASSP 94*, pp.329-332, 1994.
- [9] H. P. Robertson, *Phys. Rev.* 34, 1929.
- [10] E. U. Condon, "Remarks on uncertainty principles," *Physical Review* 34, pp. 163-164, 1929
- [11] J. Uffink and J. Lith-van Dis, "Thermodynamic uncertainty relations," *Foundations of Physics*, to appear.
- [12] J. R. Croca, *Fundamental Problems in Quantum Physics*, pp.73-82, Kluwer Academic Press, 1995
- [13] L. Cohen, "The Uncertainty Principle in Signal Analysis," *Proc. IEEE-SP Int. Symp. on Time-Frequency and Time-Scale Analysis*, pp. 182-185, 1994.
- [14] P. J. Loughlin and K. L. Davidson, "Positive Local Variances of Time-Frequency Distributions and Local Uncertainty", *Proceedings of the IEEE TFTS 98*, Pittsburgh, 1998.
- [15] P. M. Oliveira, *Engineer's Thesis*, Naval Postgraduate School, Monterey, USA, 1989.
- [16] J. C. Andrieux et al., "Optimum Smoothing of the Wigner-Ville Distribution," *IEEE Trans. Sig. Proc.*, ASSP-35, N° 6, June 1987.
- [17] A. W. Rihaczek, "Signal Energy Distribution in Time and Frequency," *IEEE Trans. Info. Theory*, Vol. IT-14, N° 3, May 1968.



# HIGH RESOLUTION FREQUENCY TRACKING VIA NON-NEGATIVE TIME-FREQUENCY DISTRIBUTIONS

Robert M. Nickel and William J. Williams

Department of Electrical Engineering and Computer Science, University of Michigan,  
Ann Arbor, MI 48109-2122, e-mail : robertn@eecs.umich.edu

## ABSTRACT

A new method for frequency tracking is presented. It combines high resolution time-frequency analysis with the discriminative power of a multiple signal classification (MUSIC, see [5]) technique. The proposed time-varying frequency estimator employs a signal adaptive time-frequency distribution (TFD) with suppressed interference terms. The adaptive TFD is able to resolve components that are closely spaced in time and frequency. The achieved resolution is superior to the one achieved by sliding window techniques which are hampered in joint time-frequency resolution by the uncertainty principle. Simulations show that the presented approach operates reliably in low signal-to-noise ratio environments. The proposed method improves and generalizes the method proposed in [1].

## 1. INTRODUCTION

The frequency tracking problem we want to address in this paper is furnished by the estimation of  $p$  deterministic frequency contours  $f_i(t) = \frac{1}{2\pi} \frac{d}{dt} \varphi_i(t)$  from a given noisy observation  $x(t)$ :

$$x(t) = z(t) + \sum_{i=1}^p x_i(t) \quad \text{with} \quad x_i(t) = X_i e^{j\varphi_i(t) + j\varrho_i} \quad (1)$$

The random process  $z(t)$  is assumed to be zero mean complex white Gaussian noise with variance  $\sigma_z^2 = E\{z(t)z^*(t)\}$ . In this section we will also assume that the phase  $\varrho_i$  of each component is an independent uniformly distributed random variable with  $\varrho_i \in [0, 2\pi]$  for  $i = 1 \dots p$ . We assume the number of components  $p$  to be known.

Optimal estimators<sup>1</sup> for  $f_i(t)$  are mathematically cumbersome and thus not practical (see [4]). Instead we will employ a suboptimal (but practical) approach based on an eigenspace analysis of the local autocorrelation function  $R_{xx}(t, \tau)$ :

$$R_{xx}(t, \tau) = E\{x^*(t - \frac{\tau}{2})x(t + \frac{\tau}{2})\} \quad (2)$$

$$= \sigma_z^2 \delta(\tau) + \sum_{i=1}^p X_i^2 e^{j[\varphi_i(t+\tau/2) - \varphi_i(t-\tau/2)]} \quad (3)$$

The general connection between the signal space formed by the  $x_i(t)$  and the eigenspace of  $R_{xx}(t, \tau)$  is very difficult to obtain. It will be shown in section 3, however, that there exists a very simple connection if we restrict ourselves to chirp components of the form:

$$\varphi_i(t) = \frac{1}{2} \alpha_i t^2 + \omega_i t \quad (4)$$

<sup>1</sup>Like a maximum likelihood estimator (MLE) for example.

Using these chirp components we can simplify the resulting local autocorrelation function  $R_{xx}(t, \tau)$ :

$$R_{xx}(t, \tau) = \sigma_z^2 \delta(\tau) + \sum_{i=1}^p X_i^2 e^{j[\alpha_i t + \omega_i]\tau} \quad (5)$$

We will describe in section 3 how to obtain estimates of the desired frequency contours  $f_i(t)$  based on an estimate of  $R_{xx}(t, \tau)$ . The next section addresses the problem of finding a proper estimate for  $R_{xx}(t, \tau)$ .

## 2. ADAPTIVE AUTOCORRELATION ESTIMATION

Finding a good estimator for  $R_{xx}(t, \tau)$  is not a trivial task. Unfortunately, the obvious choice  $\hat{R}_{xx}(t, \tau) = x^*(t - \frac{\tau}{2})x(t + \frac{\tau}{2})$  has two major drawbacks: 1) the estimate suffers from a very large variance and 2) the resulting autocorrelation at any time  $t$  is generally not positive semidefinite<sup>2</sup> in  $\tau$ . Both drawbacks can be addressed by proper smoothing. In this paper we consider a form of smoothing that takes the special structure of our signal  $x(t)$  into account.

We gain a lot of mileage by considering the cross-ambiguity function<sup>3</sup>  $A_{x_i x_k}(\theta, \tau)$  of two signal components  $x_i(t)$  and  $x_k(t)$ :

$$A_{x_i x_k}(\theta, \tau) = \int x_i^*(t - \frac{\tau}{2}) x_k(t + \frac{\tau}{2}) e^{j\theta t} dt \quad (6)$$

$$\text{with} \quad x_i(t) = X_i e^{j[\frac{1}{2} \alpha_i t^2 + \omega_i t + \varrho_i]} \quad (7)$$

Throughout the remainder of this section we will consider a specific realization of the process  $x(t)$ . This implies that each  $\varrho_i$  becomes a fixed number and  $z(t)$  becomes a fixed signal that resulted from that particular realization. The overall ambiguity function  $A_{xx}(\theta, \tau)$  of this particular  $x(t)$  can be written as<sup>4</sup>:

$$A_{xx} = A_{zz} + \sum_i A_{zx_i} + \sum_i A_{x_i z} + \sum_i A_{x_i x_i} + \sum_i \sum_{k \neq i} A_{x_i x_k} \quad (8)$$

The terms  $A_{zz}$ ,  $A_{zx_i}$ , and  $A_{x_i z}$  establish the undesired noise terms. Note that we can assume with probability one that the magnitude of each of these three terms is bounded. It is also possible to show that the cross-terms  $A_{x_i x_k}$  for  $k \neq i$  are bounded for all  $(\theta, \tau)$  if  $\alpha_i \neq \alpha_k$ . A special case arises, however, if  $\alpha_k = \alpha_i$ . Then the resulting term becomes:

$$|A_{x_i x_k}(\theta, \tau)| = \text{constant} \cdot \delta(\theta + \alpha_i \tau + \omega_k - \omega_i) \quad (9)$$

<sup>2</sup>This is obvious from the fact that the Wigner distribution, which is the Fourier transform of  $R_{xx}(t, \tau)$  in  $\tau$ , generally exhibits negative values.

<sup>3</sup>Integration is always assumed to be over  $[-\infty, +\infty]$ .

<sup>4</sup>The dependency on  $(\theta, \tau)$  is omitted. Summations are over  $1 \dots p$ .



which establishes an impulse ridge that does not pass through the origin  $(\theta, \tau) = (0, 0)$  of the ambiguity domain. In fact, the only terms that produce an impulse ridge through the origin of the ambiguity domain are the auto-components:

$$|A_{x_i x_i}(\theta, \tau)| = \text{constant} \cdot \delta(\theta + \alpha_i \tau) \quad (10)$$

We can exploit this fact in order to construct an estimator for the location of the auto-components in the ambiguity domain. Consider the following radial integral:

$$Q(\xi) = \int |A_{xx}(r \cos \xi, r \sin \xi)| e^{-r^2/2\mu^2} dr \quad (11)$$

with  $\mu$  being an arbitrary finite number. If  $A_{xx}(\theta, \tau)$  would not include the terms  $A_{x_i x_i}(\theta, \tau)$  then  $Q(\xi)$  would always attain a finite value, since all terms in (8) (except the auto-components  $A_{x_i x_i}(\theta, \tau)$ ) are bounded or have impulse ridges away from the origin. The key is that  $Q(\xi)$  becomes singular if and only if we are integrating in the direction of an impulse ridge that resulted from an auto-term. In other words, we have  $Q(\xi_i) \rightarrow \infty$  if and only if  $\alpha_i = -1/\tan \xi_i$ .

Since we will not be able to deal with infinite data sets in practical implementations of this algorithm we will observe significant spikes in  $Q(\xi)$  instead of singularities. An example can be seen in figure 3.

After getting the  $q$  spike locations  $\xi_i$  we can construct an adaptive kernel  $\phi_a(\theta, \tau)$  that supports chirp auto-terms and suppresses all other terms. Note that  $q$  denotes the number of different chirp rates in the signal  $x(t)$  and not the number of components  $p$ .

$$\phi_a(\theta, \tau) = e^{-\prod_{i=1}^q [d_i(\theta, \tau)/\sigma]^2} \quad \text{with} \quad (12)$$

$$d_i(\theta, \tau)^2 = \theta^2 + \tau^2 - (\theta \sin \xi_i + \tau \cos \xi_i)^2$$

An example for the kernel function that follows from the  $Q(\xi)$  depicted in figure 3 can be seen in figure 4. It is worth mentioning that for  $q = 2$ ,  $\xi_1 = 0$  and  $\xi_2 = \pi/2$  the function  $\phi_a(\theta, \tau)$  becomes the exponential kernel introduced by Choi and Williams in [2].

We can now construct an adaptive chirp time-frequency distribution via:

$$C_{xx}(t, \omega) = \frac{1}{4\pi^2} \iint A_{xx}(\theta, \tau) \phi_a(\theta, \tau) e^{-j\theta t - j\tau \omega} d\tau d\theta \quad (13)$$

The desired positive semidefinite estimate for the local autocorrelation function  $R_{xx}(t, \tau)$  follows by projecting  $C_{xx}(t, \omega)$  onto the set of non-negative functions in  $(t, \omega)$  from:

$$\hat{R}_{xx}(t, \tau) = \frac{1}{2} \int [C_{xx}(t, \omega) + |C_{xx}(t, \omega)|] e^{j\tau \omega} d\omega \quad (14)$$

### 3. CHIRP MUSIC

We can now use  $\hat{R}_{xx}(t, \tau)$  instead of the local autocorrelation given by equation (5) to estimate the frequency contours  $f_i(t)$ . It is beneficial for the presentation of the material in this section to switch from a continuous time description of the procedure to a discrete time description. A proper discretization of the procedures presented in the previous section is possible, but has to be omitted due to space limitations.

In discrete time we obtain the following set of equations in analogy to the equations presented in section 1:

$$x[n] = z[n] + \sum_{i=1}^p x_i[n] \quad (15)$$

$$\text{with } x_i[n] = X_i e^{j[\frac{1}{2} \alpha_i n^2 + \omega_i n + \varrho_i]} \quad (16)$$

Again,  $z[n]$  is zero mean complex white Gaussian noise with variance  $\sigma_z^2$ , and each  $\varrho_i$  is an independent random variable uniformly distributed over  $[0, 2\pi]$  for  $i = 1 \dots p$ .

$$R_{xx}[n, k] = E \{ x^*[n - \frac{k}{2}] x[n + \frac{k}{2}] \} \quad (17)$$

$$= \sigma_z^2 \delta[k] + \sum_{i=1}^p X_i^2 e^{j[\alpha_i n + \omega_i]k}$$

We can arrange  $2M + 1$  values of  $R_{xx}[n, k]$  into a  $(M + 1) \times (M + 1)$  matrix:

$$\mathbf{R}_{xx}[n] = \begin{bmatrix} R_{xx}[n, 0] & R_{xx}[n, -1] & \dots & R_{xx}[n, -M] \\ R_{xx}[n, 1] & R_{xx}[n, 0] & \dots & \vdots \\ \vdots & \vdots & \ddots & \vdots \\ R_{xx}[n, M] & \dots & \dots & R_{xx}[n, 0] \end{bmatrix}$$

A vector of signal components can be defined by

$$\mathbf{x}_i[n] = [1 \quad e^{j[\alpha_i n + \omega_i]} \quad \dots \quad e^{j[\alpha_i n + \omega_i]M}]^T \quad (18)$$

from which it is readily verified that

$$\mathbf{R}_{xx}[n] = \sum_{i=1}^p X_i^2 \mathbf{x}_i[n] \mathbf{x}_i^H[n] + \sigma_z^2 \mathbf{I} \quad (19)$$

with  $\mathbf{I}$  denoting the identity matrix. An eigenvector and eigenvalue decomposition with  $\mathbf{R}_{xx}[n] \mathbf{v}_i[n] = \lambda_i[n] \mathbf{v}_i[n]$  yields  $M + 1$  eigenvectors  $\mathbf{v}_i[n]$  and  $M + 1$  non-negative real eigenvalues  $\lambda_1[n] > \lambda_2[n] > \dots > \lambda_{M+1}[n]$ . It is readily verified that the eigenvectors  $\mathbf{v}_i[n]$  for  $i = 1 \dots p$  span the same subspace as the signal component vectors  $\mathbf{x}_i[n]$  for  $i = 1 \dots p$ . The remaining eigenvectors  $\mathbf{v}_i[n]$  for  $i = p + 1 \dots M + 1$  span a space that is orthogonal to the signal component space (see [4]). As a consequence we can construct a time-varying MUSIC estimator  $P_{xx}(n, \omega)$  via:

$$P_{xx}(n, \omega) = \frac{1}{\sum_{i=p+1}^{M+1} |e^{H(\omega)} \mathbf{v}_i[n]|^2} \quad (20)$$

$$\text{with } \mathbf{e}(\omega) = [1 \quad e^{j\omega} \quad e^{j2\omega} \quad \dots \quad e^{jM\omega}]^T \quad (21)$$

It can be shown that the locations  $(n, 2\pi f_i[n])$  of the peaks in  $P_{xx}(n, \omega)$  are precisely the desired values of the frequency contour  $f_i[n] = \alpha_i n + \omega_i$ . In a practical application the true autocorrelation matrix  $\mathbf{R}_{xx}[n]$  is replaced with an estimate obtained from a proper discretization of  $\hat{R}_{xx}(t, \tau)$  from equation (14). The next section provides an example.

### 4. EXAMPLE

The example that is considered in this section is a discrete signal  $x[n]$  which consists of four chirps buried in complex white Gaussian noise. The first chirp starts at normalized frequency

$f_{s1} = -0.04$  and chirps up to  $f_{e1} = 0.41$ . The second chirp is closely spaced to the first one from  $f_{s2} = -0.06$  to  $f_{e2} = 0.39$ . The third chirp starts at  $f_{s3} = 0.1$  and ends at  $f_{e1} = -0.45$ . Lastly, the fourth component is a stationary complex exponential with frequency  $f_4 = -0.265$ . Note that the signal has *four* components but only *three* different chirp rates. The signal-to-noise ratio for the presented case is  $SNR = 3[dB]$ . The signal is 256 samples long.

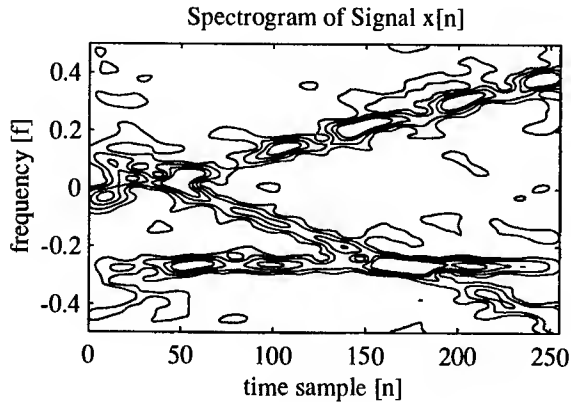


Figure 1. A spectrogram of the example signal  $x[n]$ . The employed window is a Hanning window with length 31. The window length was chosen such that one obtains an optimal tradeoff between the time and the frequency resolution.

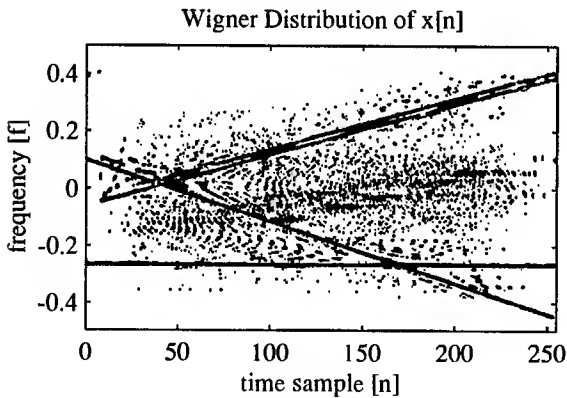


Figure 2. A Wigner distribution of the example signal  $x[n]$ . The four chirp components are clearly resolved. However, the representation is strongly distorted with cross-terms.

Figure 1 shows a spectrogram of the example signal. Two of the underlying four chirp components of the signal are reasonably well represented. The remaining two chirps however cannot be resolved since they are lying too closely spaced in time and frequency. This inherent limitation of the spectrogram carries over to any other estimation method that is based on a sliding window technique.

Figure 2 shows a Wigner distribution of the example signal. Even though the four chirps are clearly resolved it is still very difficult to use the Wigner distribution for frequency tracking purposes. Large cross-term peaks obscure the true location of the auto-terms.

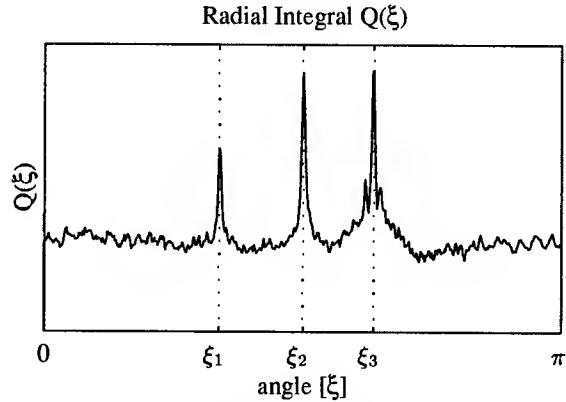


Figure 3. The radial integral  $Q(\xi)$  that results from the example signal  $x[n]$ . The three spikes that correspond to the three different chirp rates in  $x[n]$  are clearly visible.

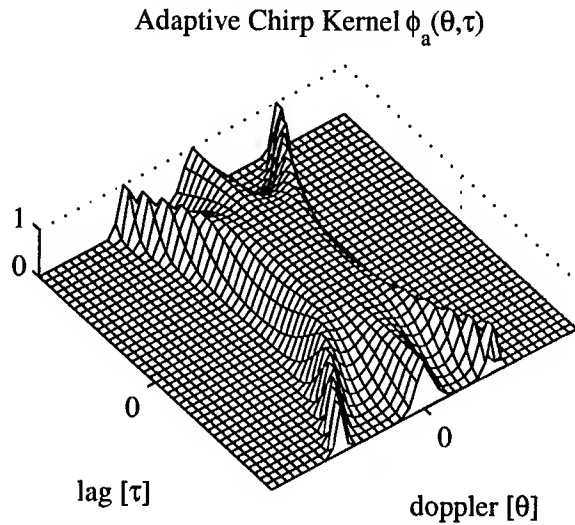


Figure 4. The center segment of the adaptive kernel  $\phi_a(\theta, \tau)$  that was obtained from the example signal  $x[n]$ . The three intersecting ridges are a consequence of the three different chirp rates in  $x[n]$ .

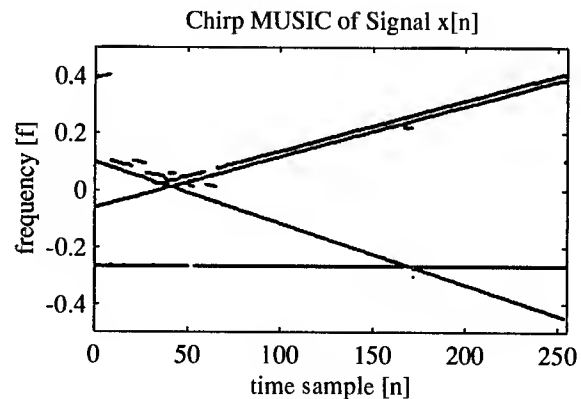


Figure 5. The adaptive time-varying MUSIC estimate  $P_{xx}(n, \omega)$  from signal  $x[n]$ . All four components are well resolved. The few visible misclassifications are due to the high noise content in the signal.

Figure 3 shows the radial integral  $Q(\xi)$  from  $A_{xx}(\theta, \tau)$  for the example signal  $x[n]$ . The three significant spikes in  $Q(\xi)$  correspond to the three different chirp rates in the signal. Figure 4 displays the center segment of the adaptive kernel function  $\phi_a(\theta, \tau)$  that followed from the given  $Q(\xi)$ . In figure 5 we can see the resulting chirp MUSIC estimate  $P_{xx}(n, \omega)$ . All four chirps are well resolved despite the low signal-to-noise ratio in the given case.

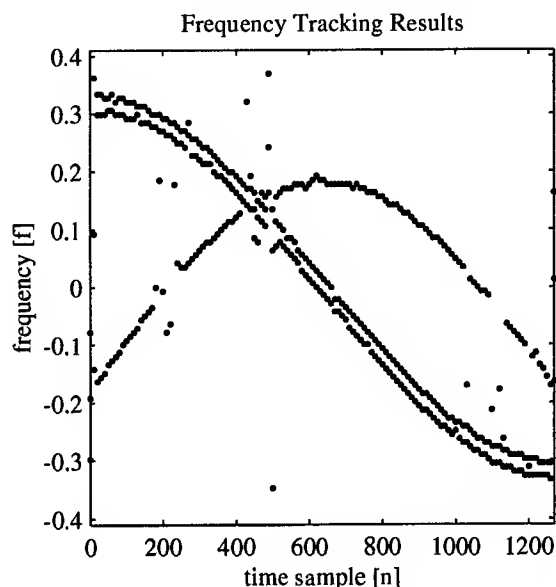


Figure 6. A simulation result for the general frequency tracking procedure described by equations (22) and (23). The underlying signal  $y[n]$  consisted of three sinusoidal frequency modulations buried in complex white Gaussian noise. The signal length was 1280 samples. The sliding window  $h[n]$  was a Hanning window with length 129. The signal-to-noise ratio was 3[dB].  $P_{yy}(m, \omega)$  was evaluated every 10 samples in time.

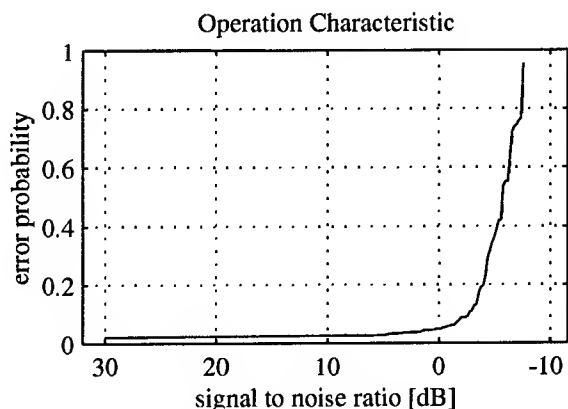


Figure 7. Error probability versus signal-to-noise ratio for the proposed algorithm. An error is defined as an event in which the estimated instantaneous frequency  $f_i[n]$  for component number  $i$  at time  $n$  is not equal to (or within the tolerance range of the spectral sampling of) the true underlying instantaneous frequency for that particular component at that particular time.

## 5. GENERAL FREQUENCY TRACKING

In the previous sections we have considered signals that were composed of linear chirp components only. It is possible to extend the proposed method to signals  $y[n]$  with arbitrary frequency contours if the individual components of the signal can be approximated locally with chirps. This is generally true for signals that have frequency contours with a small curvature. We can use a symmetric window  $h[n] = h[-n]$  with finite support ( $h[n] = 0$  for all  $|n| > N$  for some  $N$ ) to isolate the signal part that we want to approximate locally. We can track arbitrary frequency contours by using a sliding window technique according to the following equations:

$$x_m[n] = y[n - m] \cdot h[n] \quad (22)$$

The resulting estimate  $P_{yy}(m, \omega)$  is obtained from:

$$P_{yy}(m, \omega) = P_{x_m x_m}(0, \omega) \quad (23)$$

A simulation example for the proposed method is given in figure 6.

Figure 7 displays the error probability of the presented estimation algorithm. The graph resulted from a large number of simulations run with different frequency profiles and different noise levels. It is clearly visible that the method still performs well for low signal-to-noise ratios.

## 6. CONCLUSIONS

The introduced new class of adaptive high spectral resolution time-frequency distribution kernels improves and generalizes the distributions proposed in [1]. The signal dependent kernel is constructed with respect to optimal performance for signals that are composed of linear chirps in complex white Gaussian noise. Additionally, the presented time-varying multiple signal classification (MUSIC) method is used to obtain a separation between the signal subspace and the noise subspace. The provided frequency tracking simulations show that we obtain excellent estimation results even for closely spaced and rapidly changing transients. The proposed method delivers good high resolution estimates even if the underlying signal is composed of non-linear frequency contours with a low curvature. Furthermore, it is shown by simulations that the proposed method operates well in low signal-to-noise ratio cases.

## 7. REFERENCES

- [1] M. G. Amin and W. J. Williams. High spectral resolution time-frequency distribution kernels. *IEEE Transactions on Signal Processing*, 46(10):2796–2804, October 1998.
- [2] H. I. Choi and W. J. Williams. Improved time-frequency representation of multicomponent signals using exponential kernels. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, 37(6):862–871, 1989.
- [3] Leon Cohen. *Time-Frequency Analysis*. Prentice Hall, Englewood Cliffs, NJ, 1995.
- [4] Steven M. Kay. *Modern Spectral Estimation: Theory and Applications*. PTR Prentice Hall, Englewood Cliffs, New Jersey 07632, 1988.
- [5] R. O. Schmidt. *A Signal Subspace Approach to Multiple Emitter Location and Spectral Estimation*. PhD thesis, Stanford University, 1981.

# A CUMULANT SUBSPACE APPROACH TO FIR MULTIUSER CHANNEL ESTIMATION

Jing Liang and Zhi Ding

Department of Electrical and Computer Engineering, University of Iowa  
Iowa City, IA 52242  
jliang, zding@icaen.uiowa.edu

## ABSTRACT

Blind identification of FIR multiuser channels using higher-order statistics (HOS) is investigated in this work. Higher order statistical information is exploited to identify systems with equal number of users and outputs. A sufficiency condition is presented and is less stringent than many known conditions. Furthermore, a new blind identification method is developed which extends the second order subspace approach to HOS with closed-form solutions. This algorithm is shown to be capable of identifying a wider class of channels and robust to some channel order over-estimation.

## 1. INTRODUCTION

Multiple-Input Multiple-Output (MIMO) models arise in a wide range of applications such as the CDMA multiple access communications systems. There have been a number of research works dedicated to the blind identification of FIR-MIMO systems using second-order statistics (SOS) [1]-[3]. It should be noted that SOS methods are restricted to systems satisfying rather stringent conditions requiring more available outputs than inputs. In fact, even some HOS methods [4] require such conditions. Nevertheless, a notable advantage of HOS methods is that their proper use permits identifiability of a wider class of channels that may have equal number of outputs and users. Our study here focuses on this particular type of systems. Related works include [5], where blind identification and source separation conditions for MIMO system driven by colored inputs are explored. In [6], higher-order cumulant matching is used via non-linear optimization.

In this paper, we generalize our previous work [7] and propose a cumulant matrix subspace algorithm for multiuser systems with equal number of output and input signals. Unknown channel information is extracted from nullspace decomposition of several cumulant matrices which contain cross-sections of  $m$ -th order output cumulants. MIMO matrix impulse response can be identified up to a non-singular matrix. This linear HOS method is simple and admits a closed-form solution. A less stringent, sufficient identifiability condition is determined for this method. Exact knowledge of channel order is not required and only an upper bound is needed.

Supported by NSF grant CCR-9996206.

## 2. MODEL DESCRIPTION

Given a discrete  $N$ -input/ $p$ -output FIR-MIMO system, the  $i$ -th channel output signal  $x_{i,n}$  is given by

$$x_{i,n} = \sum_{u=1}^N \sum_{k=0}^q s_{n-k,u} h_{i,u}(k) + w_{i,n} \quad i = 1, \dots, p \quad (2.1)$$

Mutually independent input sequences  $\{s_{n,u}\}$  are i.i.d. non-Gaussian stationary processes with zero mean.  $h_{i,j}(n)$  is the impulse response from input  $j$  to output  $i$ . The maximum time span of  $h_{i,j}(n)$  is  $q + 1$ . Noises  $w_{i,n}$  are zero-mean stationary Gaussian processes and are independent of  $\{s_{n,u}\}$ . Let  $\tilde{x}_n = [x_{1,n} \dots x_{p,n}]^T$ ,  $\tilde{s}_n = [s_{n,1} \dots s_{n,N}]^T$ , and  $\tilde{w}_n = [w_{1,n} \dots w_{p,n}]^T$ . It then follows that  $\tilde{x}_n = \sum_{k=0}^q H_k \tilde{s}_{n-k} + \tilde{w}_n$ , where the  $p \times N$  channel response matrix  $H_n = [h_{i,j}(n)]$ .

Define a vector  $\tilde{x}[n] \triangleq [\tilde{x}_n^T, \dots, \tilde{x}_{n-L}^T]^T$ . The linear system can be described by

$$\tilde{x}[n] = H \tilde{s}[n] + \tilde{w}[n], \quad (2.2)$$

where  $\tilde{s}[n] \triangleq [\tilde{s}_n^T, \dots, \tilde{s}_{n-L-q}^T]^T$ ,  $\tilde{w}[n] \triangleq [\tilde{w}_n^T, \dots, \tilde{w}_{n-L}^T]^T$ , and the convolution matrix  $H$  is a  $(L+1)p \times (L+q+1)N$  block Toeplitz matrix

$$H = \begin{bmatrix} H_0 & H_1 & \dots & H_q & 0 & \dots & 0 \\ 0 & H_0 & H_1 & \dots & H_q & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & 0 \\ 0 & \dots & 0 & H_0 & H_1 & \dots & H_q \end{bmatrix}. \quad (2.3)$$

Denote  $\text{cum}(x_1, \dots, x_m)$  as  $m$ -th order joint cumulant of  $m$  random variables. According to [8], for  $m \geq 3$ ,

$$\begin{aligned} & \text{cum}(x_{l_1, (n-n_1)}, \dots, x_{l_m, (n-n_m)}) \\ &= \sum_{u=1}^N \gamma_{m,u} \sum_i h_{l_1,u}(i \Leftrightarrow n_1) \dots h_{l_m,u}(i \Leftrightarrow n_m), \end{aligned} \quad (2.4)$$

where  $\gamma_{m,u}$  is the  $m$ -th order kurtosis of  $s_{n,u}$ . We consider the case with symmetric signal  $s_{n,u}$  and non-zero  $\gamma_{m,u}$  for even  $m$ . Define a cumulant matrix containing cross-sections of  $m$ -th order output signal cumulant.

$$C_m^{(u,k)} \triangleq \text{cum}(\tilde{x}[n], \underbrace{\tilde{x}[n]^H, x_{l_1, (n-k)}, x_{l_1, (n-k)}^*, \dots, x_{l_m, (n-k)}^*}_{m-2 \text{ even}}) \quad (2.5)$$

From (2.2), we find that

$$C_m^{(l,k)} = H \Lambda_m^{(l,k)} H^H \quad (2.6)$$

$$\Lambda_m^{(l,k)} = \text{diag}(\underbrace{0, \dots, 0}_{k \text{ blocks}}, D_{l,0}, \dots, D_{l,q}, \underbrace{0, \dots, 0}_{(L-k) \text{ blocks}}) \quad (2.7)$$

$$D_{l,j} = \text{diag}(\gamma_{m,1} |h_{l,1}(j)|^{m-2}, \dots, \gamma_{m,N} |h_{l,N}(j)|^{m-2}). \quad (2.8)$$

Notice that  $\Lambda_m^{(l,k)}$  is block diagonal, determined by  $D_{l,j}$  ( $j = 0, 1, \dots, q$ ), each of which is an  $N \times N$  diagonal matrix. Observe that  $C_m^{(l,k)}$  and  $H$  have well-defined structures. Our algorithm take advantage of such *a priori* structural knowledge in parameter estimation. If we choose  $k = q$  and  $L \geq k + q$ , (2.6) can be rewritten as

$$C_m^{(l,k)} = H_s \Sigma^{(l,k)} H_s^H, \quad (2.9)$$

where  $H_s$  is a  $(L+1)p \times (L+1)qN$  block Toeplitz matrix

$$H_s \triangleq \begin{bmatrix} H_q & 0 & \dots & 0 \\ \vdots & H_q & \ddots & \vdots \\ H_0 & \ddots & \ddots & 0 \\ 0 & H_0 & \ddots & H_q \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & H_0 \end{bmatrix} \quad (2.10)$$

$$\Sigma^{(l,k)} = \text{diag}(\underbrace{0, \dots, 0}_{k-q \text{ blocks}}, D_{l,0}, \dots, D_{l,q}, \underbrace{0, \dots, 0}_{L-k-q \text{ blocks}}) \quad (2.11)$$

We define the  $(q+1)p \times N$  channel parameter matrix  $\mathcal{H}$  as  $\mathcal{H} \triangleq [H_q^T, \dots, H_0^T]^T$ . Channel transfer function is given by  $H(z) = \sum_{k=0}^q H_k z^{-k}$ .

### 3. IDENTIFIABILITY CONDITIONS

Most existing SOS blind identification algorithms require the convolutional matrix  $H$  to be full column rank. Consequently, it is required that (a)  $H(z)$  be full column rank for all non-zero  $z$  (i.e.  $H(z)$  is irreducible); (b)  $p > N$ .

In our HOS approach, the full column rank of  $H_s$  is necessary in order to extract nullspace of  $H_s^H$  from  $C_m^{(l,k)}$ . We first establish a sufficient condition for  $H_s$  to have full column rank.

**Theorem 3.1** Consider  $H_s$  to be a  $(d+q+1)p \times (d+1)N$  block Toeplitz matrix given by (2.10) with  $d \geq 0$ .  $H_s$  is full column rank if there exists a non-zero  $z_0 \in \mathbb{C}$  (including  $\infty$ ), such that  $H(z_0)$  has full column rank.

**Proof:** It is clear that  $p \geq N$  is an implicit assumption for  $H(z_0)$  to be full rank. Define vector  $\vec{v} \triangleq [\vec{v}_d^T, \vec{v}_{d-1}^T, \dots, \vec{v}_0^T]^T$ , where  $\vec{v}_i$  are  $N \times 1$  vectors,  $\vec{v}(z) = \sum_{i=0}^d \vec{v}_i z^{-i}$ . Then

$$H_s \vec{v} = 0 \Leftrightarrow H(z) \vec{v}(z) = 0 \text{ for all } z \quad (3.1)$$

For any  $p \times N$  polynomial matrix  $H(z)$ , we can find unimodular matrices  $\{U(z), W(z)\}$ , such that

$$U(z) H(z) W(z) = \Lambda(z) \quad (3.2)$$

Matrix  $\Lambda(z)$  is the *Smith Form* of  $H(z)$ , which is uniquely determined by monic polynomials  $\{\lambda_i(z)\}$  ( $i = 1, \dots, r$ ).  $r$  is the (normal) rank of  $H(z)$  ([9], p.390, 373). If there exists a non-zero  $z_0$  such that  $H(z_0)$  is full column rank, then  $r = N$ .  $\{\lambda_i(z)\}$  obey a *division property*:  $\lambda_i(z)$  divides  $\lambda_{i+1}(z)$  for  $i = 1, \dots, r \Leftrightarrow 1$ .  $\lambda_i(z)$  is determined by  $\lambda_i(z) = \frac{\Delta_i(z)}{\Delta_{i-1}(z)}$  ( $\Delta_0(z) = 1$ ),  $\Delta_i(z)$  is the greatest common divisor of all  $i \times i$  minors of  $H(z)$ . From the *division property*, we can prove that  $\{\lambda_i(z)\}_{i=1}^N$  have at most  $qN$  distinct zeros. This means that  $\Lambda(z)$  is rank deficient at finitely many points. Thus, one can select  $d+1$  different  $z_s$ , such that  $\Lambda(z_s)$  is full column rank. Since a unimodular matrix is nonsingular for all  $z$ , we have

$$\Lambda(z_s) W^{-1}(z_s) \vec{v}(z_s) = 0 \Leftrightarrow \vec{v}(z_s) = 0 \quad (3.3)$$

for all  $z_s$ . Write this relation as a set of linear equations and we can show that  $\vec{v} = 0$ . In other words,  $H_s \vec{v} = 0$  implies that  $\vec{v} = 0$ .  $H_s$  is thus full column rank. ■

In order for  $H_s$  to be full rank, it is certainly necessary that  $\mathcal{H}$  also has full rank. If an FIR-MIMO channel satisfies this sufficient condition,  $\mathcal{H}$  can be uniquely identified up to an ambiguity matrix  $Q$  as stated in Theorem 3.2.

**Theorem 3.2** Suppose there exists a non-zero  $z_0$  (including  $\infty$ ) such that  $H(z_0)$  has full column rank. Let  $G_s$  be a block Toeplitz matrix with the same structure and dimension as  $H_s$ .  $G$  is the channel parameter matrix of  $G_s$ . Then  $G$  is full column rank and  $\text{Range}(G_s) \subseteq \text{Range}(H_s)$  iff  $G_s = H_s A$ , where  $A$  is a block diagonal matrix given by  $A = \text{diag}(Q, \dots, Q)$  with  $Q$  to be an  $N \times N$  non-singular matrix.

**Proof:** The sufficient part is rather obvious. Now we consider the necessary part. If  $\text{Range}(G_s) \subseteq \text{Range}(H_s)$ , then there exists a square matrix  $A$  such that  $G_s = H_s A$ .  $A = [A_{ij}]$  ( $i = d, \dots, 0$ ) ( $j = 0, \dots, d$ ), where  $A_{ij}$  is  $N \times N$  square matrix. Define  $A_j(z) = \sum_{i=0}^d A_{ij} z^{-i}$ ,  $G(z) = \sum_{i=0}^q G_i z^{-i}$ . Taking the advantage of the "shift invariant" feature of Toeplitz matrices  $G_s$  and  $H_s$ , the relation  $G_s = H_s A$  is conveniently rewritten as polynomial matrix equations  $H(z) \cdot A_j(z) = G(z) \cdot z^{-(d-j)}$  for all  $z$  and  $j = 0, \dots, d$ . Equivalently, we have

$$H(z) \cdot [A_j(z) \cdot z^{d-j} \Leftrightarrow A_d(z)] = 0_{p \times N} \quad (3.4)$$

for  $j < d$ . Let  $E$  denote the set of points where  $H(z)$  is full rank. By our assumption,  $E$  includes all non-zero complex numbers except at most  $qN$  distinct elements. Thus,  $A_j(z) \Leftrightarrow A_d(z) \cdot z^{-(d-j)} = 0_{p \times N}$  for all  $z \in E$  and  $j < d$ . Naturally, this relation can be reduced to a set of polynomial matrix equations

$$\sum_{k=0}^{d-j-1} A_{kj} z^{-k} + \sum_{k=0}^j (A_{(d-j+k)j} \Leftrightarrow A_{kd}) z^{-(d-j+k)} + \sum_{k=j+1}^d (\Leftrightarrow A_{kd}) z^{-(d-j+k)} = 0, \quad j < d. \quad (3.5)$$

These polynomial equations hold true for all  $z \in E$  if and only if their coefficients are identically zero. Based on this

result, we summarize the relationship between  $A_{kj}$  and  $A_{kd}$  as follows: (1)  $A_{kj} = 0$ , for  $k = 0, 1, \dots, (d \Leftrightarrow j \Leftrightarrow 1)$ , (2)  $A_{(d-j+k)j} = A_{kd}$ , for  $k = 0, 1, \dots, j$ , (3)  $A_{(j+1)d} = A_{(j+2)d} = \dots = A_{dd} = 0$ . Applying this relation for  $j = 0 \dots, d \Leftrightarrow 1$ , it is easy to verify that  $A$  is a block diagonal matrix with same diagonal blocks  $A_{(d-j)j}$  ( $j = 0, \dots, d$ ).  $A$  can be written as  $A = \text{diag}(Q, \dots, Q)$  with  $Q$  to be  $N \times N$  square matrix. Thus,  $G = \mathcal{H}Q$ . Since  $G$  and  $\mathcal{H}$  have full column rank,  $Q$  must be non-singular. ■

#### 4. MIMO CUMULANT SUBSPACE (MCS) ALGORITHM

Theorem 3.2 establishes the fundamentals of the MCS algorithm. If the nullspace of  $H_s^H$  is estimated,  $H_s$  can be uniquely identified by solving linear equation (4.1).

$$N_s^H \hat{H}_s = 0, \quad \text{where } N_s = \text{Null}(H_s^H) \quad (4.1)$$

$\hat{H}_s$  is a block Toeplitz matrix with the same structure and dimension as  $H_s$ .

Our approach needs to extract  $N_s$  from cumulant matrix  $C_m^{(l,k)}$ . In order for this subspace method to work,  $H_s$  must be full rank, or as we have established: (1)  $p \geq N$ . (2)  $H(z_0)$  is full column rank for at least one non-zero  $z_0$ . However, equation (2.11) reveals that  $\Sigma^{(l,k)}$  is rank deficient when subchannels have different channel order or some channel coefficients are zero, even if we select  $L = 2q$ . In this case,  $\text{Null}(C_m^{(l,k)}) \supset N_s$ . Without the knowledge of the positions of zero channel coefficients and the channel order of each user, it is quite difficult to determine the dimension of  $\text{Null}(C_m^{(l,k)})$  and to make assumptions about the relationship between  $\text{Null}(C_m^{(l,k)})$  and  $N_s$ .

As a result, we propose two partial nullspace algorithms using cumulant matrices. Denote  $H_s(i)$  to be the subspace of  $H_s$  taking the first  $i$  block columns. In these two methods, we obtain the partial nullspace of  $H_s^H(1)$  (or  $H_s^H(K)$ ) from the nullspace of cumulant matrices to generate the estimate of channel parameter matrix  $\mathcal{H}$ . Recall that our definition of  $H_s$  in Theorem 3.1 is a block Toeplitz matrix with adjustable size of  $(d+q+1)p \times (d+1)N$  ( $d \geq 0$ ). Thus, Theorem 3.1 and 3.2 can be adapted to current cases without modification.

Here we assume that there exists at least one non-zero element in each column of  $H_0$ . This assumption is easily met by re-indexing time  $n$  in  $\{s_{n,u}\}$  for different sources. It helps a lot in the development of our algorithms. We first present our algorithms assuming channel order  $q$  is exactly known and consider channel order over-estimation later.

##### 4.1. Single-lag Partial Nullspace Method (SLP)

In this method, we use  $p$  cumulant matrices  $C_m^{(l,k)}$  with a single delay lag  $k$  such that  $k \geq q$  and  $L \geq k+q$ . Selecting  $k = q$  without loss of generality, we stack these cumulant matrices together and obtain their common nullspace. Define

$$C_m(k) \triangleq \begin{bmatrix} C_m^{(1,k)} \\ \vdots \\ C_m^{(p,k)} \end{bmatrix} = M_s \Sigma(k) H_s^H(q+1) \quad (4.2)$$

$$M_s = \begin{bmatrix} H_s(q+1) & & 0 \\ & \ddots & \\ 0 & & H_s(q+1) \end{bmatrix} \quad (4.3)$$

$$\Sigma(k) = \begin{bmatrix} \Sigma^{(1,k)} \\ \vdots \\ \Sigma^{(p,k)} \end{bmatrix}, \quad \Sigma^{(l,k)} = \text{diag}(D_{l,0}, \dots, D_{l,q}) \quad (4.4)$$

Here we can re-write  $\Sigma^{(l,k)}$  by removing zero blocks.  $M_s$  is full column rank as  $H_s(q+1)$  is full rank.  $\text{Null}(C_m(k)) \supseteq \text{Null}(H_s^H(q+1))$  since  $\Sigma(k)$  is possibly rank deficient. However, under the non-zero column assumption on  $H_0$ , the first  $N$  columns of matrix  $\Sigma(k)$  are full rank and independent of other columns. Thus,  $\text{Null}(C_m(k))$  must be orthogonal to  $H_s(1)$ . Although we do not know the dimension of  $\text{Null}(C_m(k))$ , a lower bound is  $d_s = (L+1)p \Leftrightarrow (q+1)N$ . Define  $N_c(1)$  to be the subspace of  $\text{Null}(C_m(k))$  corresponding to the  $d_s$  smallest eigenvalues of  $C_m(k)$ . It follows that

$$N_c(1)^H H_s(1) = 0 \quad (4.5)$$

Evidently,  $N_c(1)$  only contains partial information of the entire nullspace since  $\text{Null}(H_s^H(1))$  has dimension of  $(L+1)p \Leftrightarrow N$ .

Before moving on, we propose a simplified single-lag method for even order cumulant. Instead of stacking  $p$  cumulant matrices, we define

$$S_m(k) = \sum_{l=1}^p C_m^{(l,k)} = H_s(q+1), (k) H_s^H(q+1) \quad (4.6)$$

where  $,(k) = \sum_{l=1}^p \Sigma^{(l,k)}$ . Since there is no cancellation in the summation of entries for even  $m$ , the first  $N$  columns of  $,(k)$  is full rank. Thus, we follow the same idea presented above. In this case, relationship (4.5) and the value of  $d_s$  remain the same except that  $N_c(1)$  is estimated from  $\text{Null}(S_m(k))$ .

These two methods are labeled as SLP and SLPS respectively based on the use of  $C_m(k)$  or  $S_m(k)$  to obtain the partial nullspace  $N_c(1)$ .

$\hat{H}$  is the solution of over-determined linear equation (4.5) when window length  $L$  is large enough. To allow the unique solution, the first  $(q+1)p$  rows of  $N_c(1)$  need to have rank of no less than  $(q+1)p \Leftrightarrow N$ . Such requirement can not be guaranteed in all the cases when only single-lag cumulant matrices are utilized. By introducing multiple-lag cumulant matrices, we are able to collect more statistical information and reduce the probability for pathological cases in which (4.5) is under-determined.

##### 4.2. Multiple-lag Partial Nullspace Method (MLP)

We stack matrices  $C_m(k)$  with multiple delay lags  $k = k_1, k_1+1, \dots, k_2$  and estimate their common nullspace. The choice of delay lags and  $L$  should satisfy conditions:  $k_1 = q$ ,  $k_2 > k_1$ ,  $L = k_2 + q$ . Denote  $K$  to be the number of different delay lags used,  $K = k_2 \Leftrightarrow k_1 + 1$ . Define

$$C_m(k_1, k_2) \triangleq \begin{bmatrix} C_m(k_1) \\ \vdots \\ C_m(k_2) \end{bmatrix} = M \Sigma(k_1, k_2) H_s^H \quad (4.7)$$

$$M = \begin{bmatrix} M_m & & 0 \\ & \ddots & \\ 0 & & M_m \end{bmatrix} \quad (4.8)$$

$$\Sigma(k_1, k_2) = \begin{bmatrix} \Sigma(k_1) \\ \vdots \\ \Sigma(k_2) \end{bmatrix} \quad (4.9)$$

$$\Sigma(k) = ( \underbrace{0 \cdots 0}_{k-q \text{ blocks}} \ D_0^k \cdots D_q^k \ \underbrace{0 \cdots 0}_{K+q-k-1 \text{ blocks}} ) \quad (4.10)$$

$M_m$  is defined by replacing  $H_s(q+1)$  with  $H_s$  in (4.3). Re-write  $\Sigma(k)$  in terms of its block columns, where  $D_i^k$  ( $i = 0, \dots, q$ ) contains diagonal matrices  $\{D_{1,i}, \dots, D_{p,i}\}$ . Since  $D_0^k$  is full rank for each  $k$ , it is easy to show that the first  $KN$  columns of matrix  $\Sigma(k_1, k_2)$  is full rank. Then,  $\text{Null}(C_m(k_1, k_2))$  is orthogonal to  $H_s(K)$ . As in SLP method, we pick a subspace  $N_c(K)$  of  $\text{Null}(C_m(k_1, k_2))$  associated with the  $d_m$  smallest eigenvalues, with  $d_m = (L+1)p \Leftrightarrow (K+q)N$ . So,

$$N_c(K)^H H_s(K) = 0 \quad (4.11)$$

Again,  $N_c(K)$  is a partial nullspace of  $H_s^H(K)$ .

To obtain the estimate of  $\mathcal{H}$ , we solve linear equation (4.11) as in [1]. Define vector  $\tilde{v} \triangleq [\tilde{u}_0^T, \tilde{u}_1^T, \dots, \tilde{u}_{L-1}^T]^T$ , where  $\tilde{u}_i$ 's are  $p \times 1$  vectors. Then,

$$\tilde{v}^H H_s(K) = 0 \Leftrightarrow \text{Hank}(\tilde{v})^H \mathcal{H} = 0 \quad (4.12)$$

$$\text{Hank}(\tilde{v}) = \begin{bmatrix} \tilde{u}_0 & \tilde{u}_1 & \cdots & \tilde{u}_{K-1} \\ \vdots & \vdots & \vdots & \vdots \\ \tilde{u}_q & \tilde{u}_{q+1} & \cdots & \tilde{u}_{K+q-1} \end{bmatrix} \quad (4.13)$$

We take the solution that minimizes  $\sum_{i=1}^{d_m} \|\text{Hank}(\tilde{v}_i)^H \mathcal{H}\|_F^2$ , subject to constraint  $\mathcal{H}^H \mathcal{H} = I$ .  $\|\cdot\|_F$  stands for the Frobenius norm. Estimate  $\hat{\mathcal{H}}$  is the unit-norm eigenvectors associated with the smallest  $N$  eigenvalues of matrix  $G$ , where  $G = \sum_{i=1}^{d_m} \text{Hank}(\tilde{v}_i) \text{Hank}(\tilde{v}_i)^H$ . This also guarantee the full column rankness of  $\hat{\mathcal{H}}$ .

Similar derivation also applies for  $S_m(k)$ . We label them as MLPC and MLPS, respectively.

#### Remarks:

Note that our nullspace methods could encounter pathological cases when partial nullspace lead to non-unique solutions of linear equations. The probability of such occurrence, however, is zero. Our approach in using partial subspace has been developed for second order statistics in [1] and was further investigated in [10]. To guarantee unimodality, we have developed a method to retrieve the full nullspace  $N_s$  under strong conditions that channel order of each user is known and matrix  $H_0$  has enough non-zero rows, we will not present the details here because of the page limit.

### 4.3. Channel Order Overestimation

In practice, true channel order  $q_0$  is usually unknown and we have the knowledge of its upper bound  $q$ ,  $q \geq q_0$ . Both Theorem 3.1 and Theorem 3.2 still hold true in this situation.  $N_c(1)$  and  $N_c(K)$  obtained have smaller dimension. The Toeplitz structure of  $H_s(1)$  and  $H_s(K)$  is maintained though  $H_{q_0+1}, \dots, H_q$  are zero matrices. However, MCS

algorithms still operate properly with parameter  $q$ . The estimation performance will be degraded because generally  $\hat{H}_{q_0+1}, \dots, \hat{H}_q$  are non-zero. However, the MCS algorithms will not "collapse" when channel order is over-estimated.

## 5. SIMULATION EXAMPLES

Now we present simulation results to illustrate the performance of our algorithms. The channel inputs are two mutually independent i.i.d. QPSK signals. Mutually independent zero-mean complex white Gaussian noise is added to each output. Noisy output signals have the same SNR. We use MLPS method to estimate the unknown parameters and choose  $k_1 = q$ ,  $k_2 = 2q$  and  $L = 3q$ . Only 4-th order cumulants are used. The performance is measured by the Overall Normalized Mean Square Error (ONMSE). It is obtained by averaging NMSE of all subchannels. Results are averaged over 200 Monte Carlo runs.

$$NMSE_{ij} \triangleq \frac{\sum_{n=0}^q |h_{ij}(n) \Leftrightarrow \hat{h}_{ij}(n)|^2}{\sum_{n=0}^q |h_{ij}(n)|^2} \quad (5.1)$$

$$ONMSE \triangleq \frac{1}{pN} \sum_{i=1}^p \sum_{j=1}^N NMSE_{ij} \quad (5.2)$$

*Example 1.* We consider a 2-input/2-output system with transfer function  $H(z)$  as

$$\begin{bmatrix} 1 + 0.5z^{-1} \Leftrightarrow 0.5z^{-2} \Leftrightarrow z^{-3} & 1.6 \Leftrightarrow 0.64z^{-1} + 0.388z^{-2} \\ 0.4 + 0.6z^{-1} \Leftrightarrow z^{-2} & 0.7263z^{-1} \Leftrightarrow 0.9078z^{-2} \end{bmatrix} \quad (5.3)$$

Since  $H_{11}(z)$  and  $H_{21}(z)$  have a common zero  $z_0 = 1$ ,  $H(z)$  is not irreducible. But  $H(z)$  satisfies the identifiability conditions of our algorithms. In Figure 1, we show the performance of MLPS method at different SNR levels. As SNR increases, performance improvement is evident if more data samples are used in estimation. This example verifies the identifiability condition stated in Theorem 3.1. It also demonstrates the performance of the cumulant algorithms for ill-conditioned channels even with the partial knowledge of nullspace.

*Example 2.* In this example, we consider a case when channel order is over-estimated. The channel is a 2-input/2-output system with  $H(z)$  given by

$$\begin{bmatrix} 0.7 + z^{-1} + 0.7z^{-2} & 1.4 \Leftrightarrow 1.82z^{-1} + 0.6593z^{-2} \\ 2.7 \Leftrightarrow 0.8z^{-2} & 0.5 + 1.2z^{-1} + 0.7426z^{-2} \end{bmatrix} \quad (5.4)$$

Figure 2 shows the performance of MLPS method when channel order is exactly known. Compared with example 1, reliable estimates are generated with much smaller number of data samples for this ordinary channel. In Figure 3, we show the results when channel order  $q$  is over-estimated by 1 and 2 for 6400 data samples. Clearly, the performance degradation is mild and non-abrupt as the order difference  $q \Leftrightarrow q_0$  increases. Under channel order over-estimation, these results establish the robustness of the MCS algorithm.

## 6. CONCLUSIONS

In this paper, a new linear HOS approach for blind identification of FIR-MIMO channels is presented. Our algorithms are based on nullspace decomposition of multiple cumulant matrices. A sufficient identifiability condition for this approach is derived. These partial nullspace algorithms are capable of identifying a wide class of channels including ones not identifiable due to ill-conditioning for some existing methods. Finally, our approach is less sensitive and more robust to channel order over-estimation.

## REFERENCES

- [1] E.Moulines, P.Duhamel, J.F.Cardoso, and S.Mayrargue, "Subspace methods for the blind identification of multichannel FIR filters," *IEEE Trans. on Signal Processing*, vol. 43, no.2, pp. 516-525, Feb 1995.
- [2] Ph.Loubaton and E.Moulines, "Application of blind second order statistics MIMO identification methods to the blind CDMA forward link channel estimation," *ICASSP'99*, vol. 5, pp. 2543-2546, 1999.
- [3] K.Abed-Meraim, P.Loubaton, and E.Moulines, "A subspace algorithm for certain blind identification problems," *IEEE Trans. on Info. Theory*, vol. 43, no.2, pp. 499-511, March 1997.
- [4] M.Martone, "Subspace methods for blind identification of multichannel FIR filters using space-time contraction of cumulants," *IEEE Signal Processing letters*, vol. 5, no.7, pp. 180-184, July 1998.
- [5] Y.Inouye and K.Hirano, "Cumulant-based blind identification of linear multi-input-multiple-output system driven by colored inputs," *IEEE Trans. on Signal Processing*, vol. 45, no.6, pp. 1543-1552, June 1997.
- [6] J.K.Tugnait, "Identification and deconvolution of multichannel linear non-Gaussian processes using higher order statistics and inverse filter criteria," *IEEE Trans. on Signal Processing*, vol. 45, no.3, pp. 658-672, Mar. 1997.
- [7] Z.Ding, "A cumulant matrix subspace algorithm for blind single FIR channel identification," *Proc. IEEE Signal Processing Workshop on Higher-Order Statistics*, pp. 85-88, Caesarea, Israel, June 1999.
- [8] D.R.Brillinger, *Time Series: Data Analysis and Theory*, McGraw-Hill Inc., 1981.
- [9] T.Kailath, *Linear System*, Englewood Cliffs, NJ, Prentice-Hall, 1980.
- [10] K.Abed-Meraim and Yingbo Hua, "Blind identification of multi-input multi-output system using minimum noise subspace," *IEEE Trans. on Signal Processing*, vol. 45, no.1, pp. 254-258, Jan. 1997.

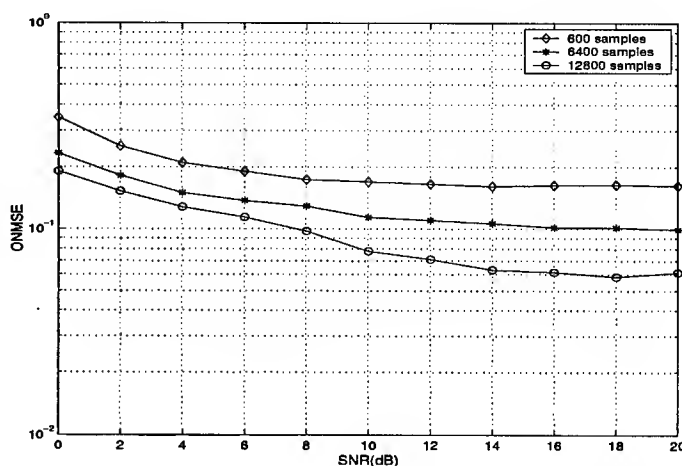


Figure 1: Example 1: Performance of MLPS at different SNR levels

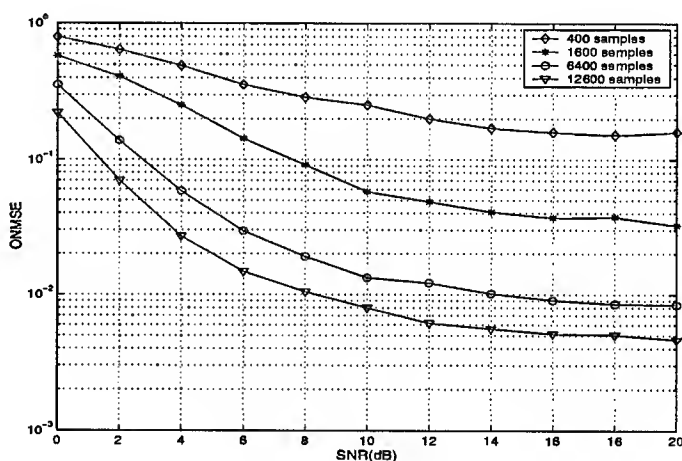


Figure 2: Example 2: Performance of MLPS at different SNR levels

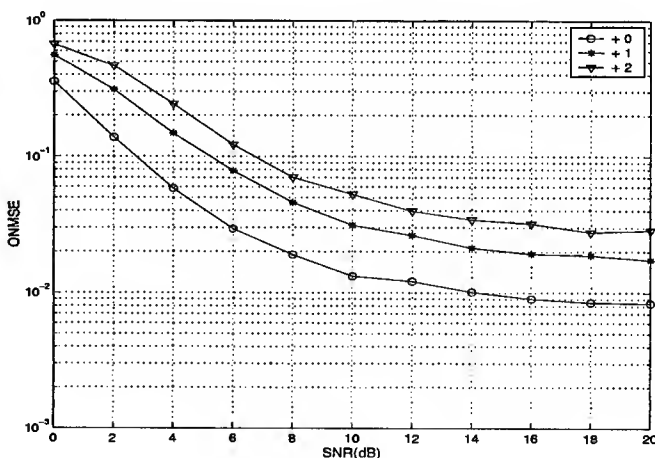


Figure 3: Example 2: Performance comparison for channel order over-estimation with 6400 samples



# AN EFFICIENT FORTH ORDER SYSTEM IDENTIFICATION (FOSI) ALGORITHM UTILIZING THE JOINT DIAGONALIZATION PROCEDURE

Adel Belouchrani

Belkacem Derras

Electrical Engineering Department  
Ecole Nationale Polytechnique  
P.O. Box 182 El-Harrach 16200, Algiers, Algeria  
belouchrani@hotmail.com

Cirrus Logic Inc.  
305 Interlocken Parkway  
Broomfield CO 80021, USA  
bderass@colorado.cirrus.com

## ABSTRACT

This paper introduces a new linear algebraic method for blind identification of a nonminimum phase FIR system. The proposed approach relies only on stationary fourth order statistics and is based on the 'joint diagonalization' of a set of fourth-order cumulant matrices. Its performance is illustrated via some numerical examples. Further this method turns out to overcome the problem of having some zero taps in the system impulse response.

## 1. INTRODUCTION

The blind identification problem of linear system is of great interest in diverse fields including speech processing, data communication, and geophysical or biological data processing due to its relevance to blind deconvolution. The general purpose of blind identification is to identify an unknown system driven by unobservable input based only on its output. This paper deals with a particular case which is the identification of a finite impulse response (FIR) system driven by a non-Gaussian white input. It is well known that second-order statistics are sufficient to identify any minimum phase system [1] and that a nonminimum phase system cannot be uniquely identified using only second-order statistics. On the other hand, it is shown for a non-Gaussian input that consistent estimates of the parameters of any FIR system can be obtained by using higher-order statistics or cumulants of the observed data [2]. This is because the higher-order statistics preserve the phase characteristics (up to a linear phase shift), unlike the second-order statistics.

Several solutions exist for the identification of the non-minimum phase FIR system using high-order statistics. Some of them are closed-form solutions [3], others are linear algebraic methods [4], or nonlinear optimization approaches [5]. The linear algebraic approaches do not perform as well as the nonlinear optimization methods but they are computationally attractive and can be used as good initial guess for the nonlinear optimization methods that usually suffer from local minima and ill convergence.

Recently, a new algebra tool referred to as 'Joint Diagonalization' has been successfully applied to some signal processing problems such as blind source separation [6, 8, 9], blind identification of linear-quadratic models [10] and source localization [11, 12, 13]. Herein, our aim is to apply this tool to the blind identification of FIR systems.

Hence, we propose a new linear algebraic approach based on a *joint-diagonalization* of a set of fourth-order cumulant matrices. The recovery (identification) of the system impulse response from this process is made possible by the existing relationships between its taps and those cumulants. Moreover, the identification procedure described herein utilizes the sample fourth-order cumulant matrices which are estimated from a finite sequence of the (possibly noisy) output signal.

## 2. PROBLEM FORMULATION

Consider a linear time invariant (LTI) FIR system described by

$$y(n) = \sum_{i=0}^q h(i)s(n-i) + v(n) \quad (1)$$

where

- $y(n)$  output sequence,
- $h(n)$  impulse response of the FIR LTI system that is allowed to be nonminimum phase,
- $s(n)$  input to the system,
- $v(n)$  additive noise.

The assumptions made about the data model are as follows.

- A1) The input process  $\{s(n)\}$  is an i.i.d. zero-mean non-Gaussian stationary process.
- A2) The noise process  $\{v(n)\}$  is a Gaussian, perhaps colored, zero-mean stationary process independent of  $\{s(n)\}$ .
- A3)  $h(n) = 0, n < 0, h(n) = 0, n > q, h(q) \neq 0$ , and  $h(0) = 1$ , which fixes the inherent scale ambiguity.

Our task in this paper is to identify the parameters of the FIR system  $(h(n), n = 1, \dots, q)$  using the fourth order cumulants of the measured output process  $y(n)$  and assuming that the FIR system order  $q$  is known. Our motivation in using the fourth order cumulants for the blind identification problem under assumptions A1) and A2) comes from the fact that Gaussian processes have identically zero cumulants of any order greater than two. Moreover, third-order cumulants vanish even for non-Gaussian random processes with symmetric distributions, but fourth order cumulants generally do not for practically useful random processes.

Hence, fourth order cumulants have two advantages; one is that the additive Gaussian noises do not affect 4-th order cumulants, and the other is that 4-th order cumulants can be used in many more situations than third order cumulants.

### 3. THE FOURTH-ORDER CUMULANT MATRICES

Let's start by giving the definition of the fourth order cumulants of a stationary random process.

**Definition 1** Let  $y(n)$  be a stationary random process. Then, the fourth-order cumulant of  $y(n)$   $c_y(k_1, k_2, k_3)$  for fixed integers  $k_1, k_2$  and  $k_3$  is defined by

$$c_y(k_1, k_2, k_3) = \text{cum}[y(n), y(n+k_1), y(n+k_2), y(n+k_3)]$$

In particular, when  $y(n)$  is zero-mean, we have the following expression,

$$\begin{aligned} c_y(k_1, k_2, k_3) &= E[y(n)y(n+k_1)y(n+k_2)y(n+k_3)] \\ &\quad - E[y(n)y(n+k_1)]E[y(n+k_2)y(n+k_3)] \\ &\quad - E[y(n)y(n+k_2)]E[y(n+k_1)y(n+k_3)] \\ &\quad - E[y(n)y(n+k_3)]E[y(n+k_1)y(n+k_2)] \end{aligned} \quad (2)$$

Now, we define the fourth-order cumulant matrices.

**Definition 2** Let  $y(n)$  be a stationary random process. Then, the  $(2q+1) \times (2q+1)$  fourth-order cumulant matrix of  $y(n)$   $S_y(k)$  for a fixed integer  $k$  is defined by

$$S_y(k) = \begin{bmatrix} c_y(-q, -q, k) & \cdots & c_y(-q, 0, k) & \cdots & c_y(-q, q, k) \\ \vdots & & \vdots & & \vdots \\ c_y(0, -q, k) & \cdots & c_y(0, 0, k) & \cdots & c_y(0, q, k) \\ \vdots & & \vdots & & \vdots \\ c_y(q, -q, k) & \cdots & c_y(q, 0, k) & \cdots & c_y(q, q, k) \end{bmatrix} \quad (3)$$

where  $c_y(i, j, k)$ ,  $-q \leq i, j \leq q$ , are the 4-th order cumulants of  $y(n)$ .

Next, we examine the relations between the cumulants of inputs and those of outputs. According to (1) and assumptions A1) and A2), the 4-th order cumulant of the output system is given by

$$c_y(i, j, k) = \gamma_s \sum_{l=0}^q h(l+i)h(l+j)h(l+k) \quad (4)$$

where  $\gamma_s = c_s(0, 0, 0)$ . In the above relation, we have taken into account the fact that the 4-th order cumulant of the additive Gaussian noise vanishes. Replacing (4) into (3) and after some algebraic manipulations, the 4-th order cumulant matrices show to have the following structure:

$$S_y(k) = \text{sign}(\gamma_s) H \Delta(k) H^T \quad (5)$$

where the superscript  $T$  designs the transpose,

$$H = \begin{bmatrix} 0 & \cdots & 1 [h(q)\sqrt{|\gamma_s|}] \\ \vdots & & \vdots \\ 0 & \cdots & h(q-1) [h(q)\sqrt{|\gamma_s|}] \\ 1 [h(0)\sqrt{|\gamma_s|}] & \cdots & h(q) [h(q)\sqrt{|\gamma_s|}] \\ \vdots & & \vdots \\ h(q-1) [h(0)\sqrt{|\gamma_s|}] & \cdots & 0 \\ h(q) [h(0)\sqrt{|\gamma_s|}] & \cdots & 0 \end{bmatrix} \quad (6)$$

and  $\Delta(k) = \text{diag}[\frac{h(k)}{1}, \frac{h(k+1)}{h(1)}, \dots, \frac{h(k+q)}{h(q)}]$ , for  $k = -q, \dots, q$ . Note that  $H$  is a  $(2q+1) \times m$  matrix and  $\Delta(k)$  is an  $m \times m$  matrix, where  $m$  is equal to the number of non-zero FIR taps. According to assumption A3), we do have:  $2 \leq m \leq q+1$ .

Next, we propose a new procedure for the estimation of the FIR system parameters  $(h(n), 0 \leq n \leq q)$  that exploits the nice structure (5) of the 4-th order cumulant matrices.

### 4. THE PROPOSED IDENTIFICATION APPROACH

**Orthonormalizing:** The first step of the proposed estimation procedure consists of orthonormalizing the fourth-order cumulant matrices. This is achieved using an orthonormalizing matrix  $W$ , i.e. a  $m \times (2q+1)$  matrix such that  $I = W[\text{sign}(\gamma_s)S_y(0)]W^T$ . Replacing the expression of  $\text{sign}(\gamma_s)S_y(0) = H\Delta(0)H^T = HH^T$  in the latter expression, shows that

$$I = (WH)(WH)^T \quad (7)$$

so that  $WH$  is a  $m \times m$  unitary matrix. For any whitening matrix  $W$ , it thus exists a  $m \times m$  unitary matrix  $U$  such that

$$WH = U \text{ or } H = W^\# U \quad (8)$$

where the superscript  $\#$  denotes the Moore-Penrose pseudoinverse:  $W^\# = W^T(WW^T)^{-1}$ . The orthonormalizing matrix  $W$  can be determined from the eigendecomposition of the fourth order cumulant matrix  $S_y(0)$  provided that  $S_y(0)$  is positive definite. If the kurtosis of the input data is negative,  $-S_y(0)$  should be used instead. The sign of the input data kurtosis can be deduced from the sign of the eigenvalues of  $S_y(0)$ .

**Fourth-order identification principle:** Now consider the orthonormalized fourth order cumulant matrices  $\underline{S}_y$  defined as

$$\forall k \neq 0 \quad \underline{S}_y(k) = W[\text{sign}(\gamma_s)S_y(k)]W^T. \quad (9)$$

Pinning the definition (5) and (8) into (9), it comes:

$$\begin{aligned} \forall k \neq 0 \quad \underline{S}_y(k) &= (WH)\Delta(k)(WH)^T \\ &= U\Delta(k)U^T. \end{aligned} \quad (10)$$

Since the matrix  $U$  is unitary and the matrix  $\Delta(k)$  is diagonal, the latter equation shows that any orthonormalized fourth-order cumulant matrix is diagonal in the basis of the

columns of the matrix  $U$  (the eigenvalues of  $\underline{S}_y(k)$  being the diagonal entries of  $\Delta(k)$ ).

If, for  $k \neq 0$  the diagonal elements of  $\Delta(k)$  are all distinct, the unitary matrix  $U$  may be 'uniquely' (i.e. up to permutation and phase shifts) retrieved by computing the eigendecomposition of  $\underline{S}_y(k)$ . Indeterminacy occurs in the case of degenerate eigenvalues, i.e. when  $[\Delta(k)]_{ii} = [\Delta(k)]_{jj}$ ,  $i \neq j$ .

The situation is more favorable when considering simultaneous diagonalization of the set  $\{\underline{S}_y(k) | k = -q, \dots, -1, 1, \dots, q\}$  of  $2q$  orthonormalized fourth-order cumulant matrices. This set is simultaneously diagonalizable by the unitary matrix  $U$  as in (10).

The matrix  $U$  is unique (to a permutation matrix and phase factors) if, and only if, for any pair  $(i, j)$ , there exists an integer  $k$  such that  $[\Delta(k)]_{ii} \neq [\Delta(k)]_{jj}$ . Of course, the simultaneous diagonalization holds only for the exact statistics; sample statistics may only be *approximately* simultaneously diagonalized under the same unitary transformation. This calls for the definition to an *approximate* simultaneous diagonalization.

**Joint approximate diagonalization:** The *joint approximate diagonalization* can be explained by first noting that the problem of the diagonalization of a single  $n \times n$  symmetric matrix  $M$  is equivalent to the minimization of the criterion [14]:

$$C(M, V) \stackrel{\text{def}}{=} - \sum_i |v_i^T M v_i|^2 \quad (11)$$

over the set of unitary matrices  $V = [v_1, \dots, v_n]$ . Hence, the *joint approximate diagonalization* of a set  $\mathcal{M} = \{M_k | k = 1, \dots, K\}$  of  $K$  arbitrary  $n \times n$  matrices is naturally defined as the minimization of the following criterion:

$$C(V, \mathcal{M}) \stackrel{\text{def}}{=} - \sum_k C(M_k, V) = - \sum_{ki} |v_i^* M_k v_i|^2 \quad (12)$$

under the same unitary constraint. An efficient joint approximate diagonalization algorithm can be found in [6] which is a generalization of the Jacobi technique for the exact diagonalization of a single symmetric matrix [14].

**Fourth-Order System Identification (FOSI):** We now have at hand all the necessary ingredients to derive the main identification procedure; it comprises the following steps

- From the eigendecomposition of the sample estimate of the fourth-order cumulant matrix  $\hat{S}_y(0)$ , estimate an orthonormalizing matrix  $\hat{W}$  (by computing a square-root of the pseudo-inverse of  $\hat{S}_y(0)$ ),
- Determine the unitary matrix  $\hat{U}$  by minimizing criterion (12) for the set of the orthonormalized sample 4-th order cumulant matrices  $\{\hat{S}_y(k), k = -q, \dots, -1, 1, \dots, q\}$ .
- Obtain an estimate of the matrix  $\hat{H}$  as  $\hat{H} = \hat{W}^* \hat{U}$ .
- Select the column of  $\hat{H}$  corresponding to the largest absolute value of the diagonal entries of  $\hat{\Delta}(q)$ , and save the  $q+1$  bottom elements of this column into vector  $\hat{f}_1$ .
- Obtain an estimate of the FIR system as  $\hat{h}_{(q)} = \frac{\hat{f}_1}{\hat{f}_1(1)}$ .

- From the  $q+1$  top elements of the column of  $\hat{H}$  corresponding to the largest absolute value of the diagonal entries of  $\hat{\Delta}(-q)$ , saved into vector  $\hat{f}_2$ , obtain another estimate of the FIR system as  $\hat{h}_{(-q)} = \frac{\hat{f}_2}{\hat{f}_2(1)}$ .
- A third estimate of the FIR system can be obtained by averaging the two previous estimates, i.e.  $\hat{h}_{(\text{average})} = \frac{\hat{h}_{(q)} + \hat{h}_{(-q)}}{2}$ .

The steps that provide the estimates  $\hat{h}_{(q)}$ ,  $\hat{h}_{(-q)}$  and  $\hat{h}_{(\text{average})}$  are referred to as  $FOSI_{(q)}$ ,  $FOSI_{(-q)}$  and  $FOSI_{(\text{average})}$ , respectively.

## 5. SIMULATION RESULTS

In the simulated environment, an FIR LTI system is considered. The input sequence is a zero-mean uniform binary process with unit variance. The system output is corrupted by a stationary Gaussian noise. The mean square error (MSE) of the estimated FIR system coefficients is obtained by averaging the results of 500 independent trials. All curves are labeled with the steps used for the identification process. On all the plots, the Cramer-Rao lower bound (CRLB) is provided to serve as a reference.

**Example 1:** In this example, we consider the following FIR system:  $h = [1 \ -2 \ 2 \ 4]$ . In figure 1, the MSE is plotted in dB as a function of the noise level for a sample size  $T = 50000$ . This figure shows that the performances are constant versus the noise level. This suggests that the proposed approach is robust to the measurement noise. In figure 2, the noise level is kept constant at -20 dB. The figure shows the MSE in dB plotted against the sample size. The plot evidences a significant improvement in performance by including a large sample size in the estimation of the sample fourth-order statistics.

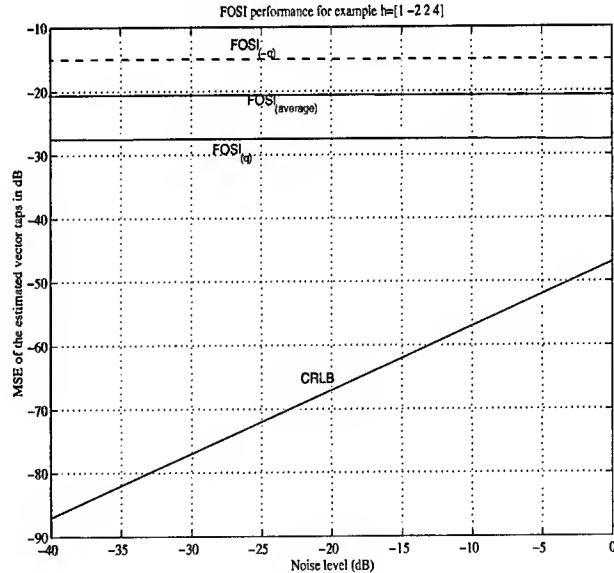


Figure 1: Mean Square Error versus noise level for example 1.

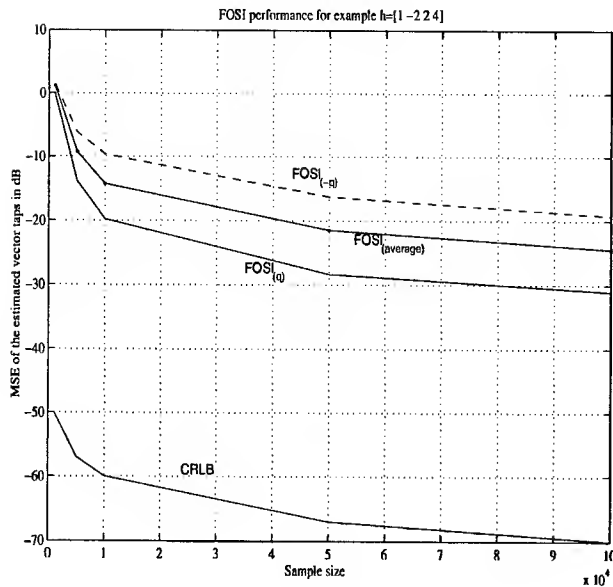


Figure 2: MSE versus sample size for example 1.

**Example 2:** In this example, we consider the following FIR system:  $h = [1 \ -2 \ 0 \ 1]$ . In figure 3, the MSE is plotted in dB as a function of the sample size for a noise level of -20 dB. This figure shows that the proposed identification method performs well when the FIR system presents zero taps.

**Example 3:** Here, we consider the following FIR system:  $h = [1 \ -1.87 \ 3.02 \ -1.435 \ 0.49 \ -0.8]$ . This example is taken from [4]. The results obtained in this section cannot be compared with those of [4] because this reference considers third-order statistics for the cumulant estimation, while we consider 4-th order statistics. Hence the comparison would not be fair. In figure 4, the MSE is plotted in dB as a function of the sample size for a noise level of -20 dB. The figure shows increase in performance allowed by a better accuracy of the sample 4-th order cumulants when including a large number of samples in the identification procedure.

Through all these examples and other extensive experiments not reported here, one notices that  $FOSI_{(q)}$  performs better than  $FOSI_{(-q)}$  for an energy of  $h(q)$  greater than the energy of  $h(0) = 1$ . When the energy of  $h(q)$  is lower than the energy of  $h(0) = 1$ , the situation is reversed, i.e.  $FOSI_{(-q)}$  performs better than  $FOSI_{(q)}$ . These facts can be explained by resorting to the expressions of  $\Delta(q)$  and  $\Delta(-q)$ .

## 6. CONCLUSION

In this paper, the problem of blind identification of FIR system based only on fourth-order statistics has been investigated. An algebraic solution based on the joint diagonalization of a set of orthonormalized 4-th order cumulant matrices has been proposed. Numerical simulations have been performed to assess the performance of the proposed method. These show robustness of the proposed approach with respect to the measurement noise. Moreover,

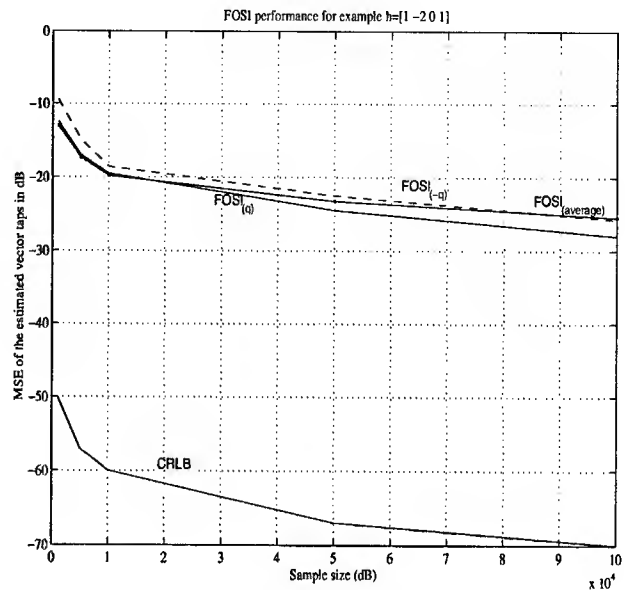


Figure 3: MSE versus sample size for example 2.

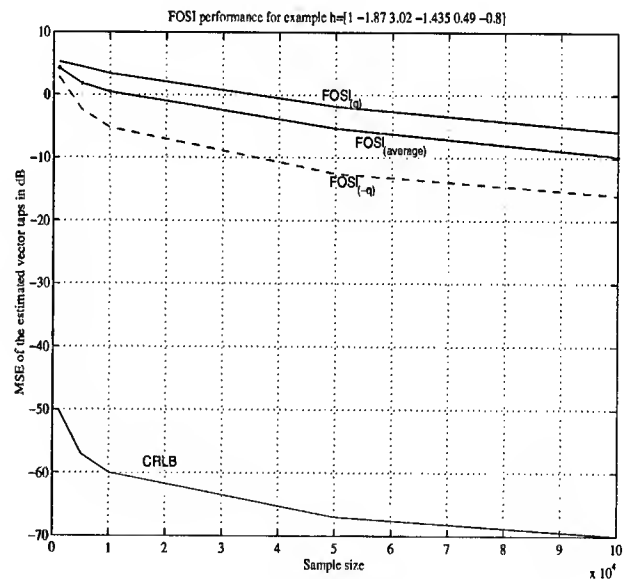


Figure 4: MSE versus sample size for example 3.

the method turns out to overcome the problem of having some zero taps in the system impulse response.

[14] G. H. Golub and C. F. V. Loan, *Matrix computations*. The Johns Hopkins University Press, 1989.

## 7. REFERENCES

- [1] J. L. Lacoume, P. O. Amblard and P. Comom, *Statistiques d'ordre supérieur pour le traitement du signal*. Masson, 1997.
- [2] D. Donoho, "On minimum entropy deconvolution," *Applied time-series analysis*, Academic Press, pp. 565-609, 1981.
- [3] A. Swami and J. M. Mendel, "Closed-form recursive estimation of Ma coefficients using autocorrelations and third-order cumulants," *IEEE Trans. on ASSP*, vol. 37, pp. 1794-1795, Nov. 1989.
- [4] L. Srinivas and K. V. S. Hari, "FIR System Identification Based on Subspaces of a Higher Order Cumulant Matrix," *IEEE Trans. on SP*, vol. 44, no. 6, pp. 1485-1491, Jun. 1996.
- [5] K. S. Lii and M. Rosenblatt, "Deconvolution and estimation of transfer function phase and coefficients of non-Gaussian linear processes," *Ann. Stat.*, vol. 10, pp. 1195-1208, 1982.
- [6] A. Belouchrani and K. Abed Meraim and J.-F Cardoso and E. Moulines, "A blind source separation technique using second order statistics," *IEEE Trans. on SP*, vol. 45, no. 2, pp. 434-444, Feb. 1997.
- [7] A. Belouchrani and M. G. Amin, "Blind Source Separation using Time-Frequency Distributions: Algorithm and Asymptotic Performance," *IEEE Proc. ICASSP*, Germany, Apr. 1997.
- [8] A. Belouchrani and M. G. Amin, "Blind Source Separation Based on Time-Frequency Signal Representation," *IEEE Trans. on SP*, vol. 46, no. 11, pp. 2888-2897, Nov. 1998.
- [9] A. Belouchrani and M. Amin, "On the Use of Spatial Time Frequency Distributions For Signal Extraction," *Special issue of the journal "Multidimensional Systems and Signal Processing"*, Kluwer Academic Publishers, Oct. 1998.
- [10] K. Abed-Meraim, A. Belouchrani and Y. Hua, "Blind Identification of a Linear-Quadratic Mixture of Independent Components Based on Joint Diagonalization Procedure," *IEEE Proc. ICASSP*, Atlanta, USA, May, 1996.
- [11] A. Belouchrani and M. G. Amin, "Time-Frequency MUSIC: A new array signal processing method based on time-frequency signal representation," *IEEE Signal Processing Letter*, vol. 6, No. 5, pp. 109-110, May 1999.
- [12] A. Belouchrani, M. G. Amin and K. Abed-Meraim, "Direction Finding in Correlated Noise Fields Based on Joint Block-Diagonalization of Spatio-Temporal Correlation Matrices," *IEEE Signal Processing Letter*, Vol. 4, No 9, pp-266-268, Sept. 1997.
- [13] A. Belouchrani, K. Abed-Meraim, and Y. Hua, "Jacobi-like Algorithms for Joint Block Diagonalization: Application to Source Localization," *In Proc. IS-PACS (Melbourne, Australia)*, Nov. 1998.

# UNITY-GAIN CUMULANT-BASED ADAPTIVE LINE ENHANCER

R. R. Gharieb<sup>†1</sup>, Y. Horita<sup>2</sup>, T. Murai<sup>2</sup> and A. Cichocki<sup>1</sup>

1 Lab. for Advanced Brain Signal Processing  
Brain Science Institute, RIKEN, Wako-Shi, Saitama, 351-0198, JAPAN  
E-mail: [reda@bsp.brain.riken.go.jp](mailto:reda@bsp.brain.riken.go.jp), Fax: + 81-48-467-9694

2 Toyama University, Faculty of Engineering  
Department of Electrical and Electronics Engineering  
3190 Gofuku, Toyama 930-8555, JAPAN

## ABSTRACT

In this paper, a unity-gain cumulant-based adaptive line enhancer (UGCBALe) is presented. This enhancer is formulated by adaptive filtering the output of the CBALE to adjust the overall gain to unity. Owing to the unity-gain feature, this enhancer can be, for example, utilized as a sinusoidal interference canceller by subtracting the enhanced output from the noisy input. The UGCBALe is insensitive to Gaussian noise, either white or colored and its performance in the case of colored non-Gaussian noise has been extensively investigated. Simulation results are presented to show the effective performance of the UGCBALe in comparison with the conventional adaptive line enhancer (ALE) when the noise is colored uniformly distributed random process (UDRP).

**Keywords-** Higher-order Statistics, non-Gaussian noise, Adaptive Line Enhancer.

## 1. INTRODUCTION

Conventional adaptive line enhancer (ALE) proposed by Widrow [1] has been successfully employed for the enhancement of sinusoidal signal in uncorrelated (white) noise. The ALE provides also cancellation of the sinusoidal signal interfering a broadband signal. The cumulant-based adaptive line enhancer (CBALE), proposed in [4], [2], is effectively capable of enhancing sinusoidal signal in correlated (colored) Gaussian noise. In spite of the fact that the CBALE outperforms the conventional ALE in colored Gaussian noise case, it provides an unknown gain, therefore its application as a sinusoidal interference canceller is limited by this unknown gain. It has been shown that the gain of the CBALE is only known in the case of a single sinusoid [4]. Therefore, in this case, a sinusoidal interference canceller is available.

In this paper, a unity-gain cumulant-based adaptive line enhancer (UGCBALe) is presented. This enhancer is a modified version of the non-unity-gain CBALE described in [4], [2]. This modification makes the presented enhancer perform as an adaptive sinusoidal interference canceller by subtracting its output from its input. Owing to recent analysis and results described in [6], [7] which have proved that employing higher-order cumulants is an effective approach to handling sinusoidal signal corrupted by additive colored non-Gaussian noise, the performance of the presented enhancer in the case of colored non-Gaussian noise is investigated. Section

2 gives background, including the signal model and a brief review of both the ALE and the non-unity-gain CBALE. Section 3 presents the novel UGCBALe. Section 4 presents illustrative simulation results, which are concerned with enhancing sinusoids in colored uniformly distributed random process (UDRP) noise. Finally, Section 5 gives the conclusion.

## 2. BACKGROUND

In this section, the signal model is described and a brief overview of both the conventional ALE and the non-unity-gain CBALE is presented.

### 2.1 The Signal Model

The observed signal  $x(n)$  is modeled as a sum of multiple sinusoids  $s(n)$  plus zero-mean additive colored noise  $v(n)$ , i.e.,

$$x(n) = s(n) + v(n) = \sum_{m=1}^P A_m \cos(2\pi f_m n + \varphi_m) + v(n) \quad (1)$$

where the amplitudes  $A_m$  and phases  $\varphi_m$  are deterministic constants. The frequencies  $0 < f_m < 0.5$  are unknown either constants or time varying parameters, and obey the constraints described in [11]. The additive noise  $v(n)$  is a zero-mean colored random process with unknown spectral density. It is assumed that  $v(n)$  is the output of a stable, linear shift-invariant (LSI) filter driven by white either Gaussian or non-Gaussian random process with bounded eighth-order moment. A local signal-to-noise ratio ( $SNR_m$ ) is defined as  $10\log_{10}(A_m^2/2\sigma_v^2)$ ,

where  $\sigma_v^2$  is the noise variance. The objective is to restore the sinusoidal signal  $s(n)$  given a single record of the observed noisy signal  $x(n)$ . If  $v(n)$  is of interest, then, another objective is to cancel the sinusoidal signal from the observed signal  $x(n)$ .

### 2.2 The ALE

Because we will compare the results of the presented UGCBALe with that of the conventional ALE, the ALE can be briefly reviewed as follows. The output of the adaptive filter working as a linear predictor is computed by

<sup>†</sup> Dr. R. R. Gharieb is on leave from the Faculty of Engineering, Assiut University, Egypt

$$y(n) = \sum_{i=0}^{2M} w_i(n)x(n-i) \quad (2)$$

The error signal is given by

$$e(n) = x(n-\Delta) - y(n) \quad (3)$$

where  $w_i(n)$  are the adaptive filter coefficients,  $x(n)$  is the reference signal,  $x(n-\Delta)$  is the primary signal and  $\Delta$  is the decorrelation delay, which is enough in the case of white noise to decorrelate noise of both reference and primary inputs. The appropriate value of  $\Delta$  is chosen equal to  $M$  in order to keep the causality of the adaptive filter. The normalized least mean square (NLMS) update equation is given by

$$w_i(n+1) = w_i(n) + \mu e(n)x(n-i)/\gamma \quad (4)$$

where  $\mu$  is a positive step size and  $\gamma = \sum_{i=0}^{2M} x^2(n-i)$ .

### 2.3 The Non-Unity-Gain CBALE

The non-unity-gain CBALE shown in Figure 1 is an FIR adaptive filter whose input is the noisy signal  $x(n)$  and output is the enhanced sinusoidal signal  $z(n)$ . That is, the output signal  $z(n)$  is given by

$$z(n) = \sum_{i=0}^{2L} h_i(n)x(n-i) \quad (5)$$

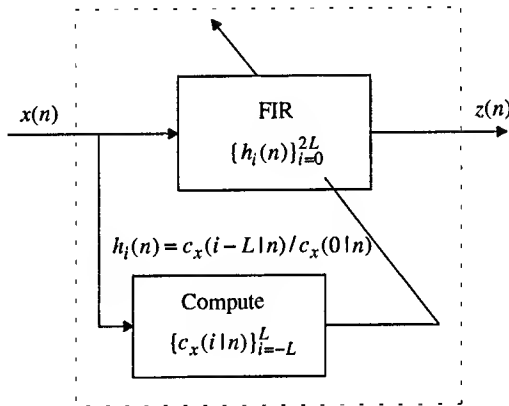


Figure 1. The cumulant-based adaptive line enhancer (CBALE).

The adaptive filter impulse response  $h_i(n)$  is computed recursively using [4], [2]

$$c_x(i|n) = \alpha_1 c_x(i|n-1) + (1-\alpha_1)x(n-i) \times [x^3(n) - 3p_x(n)x(n)], \quad |i| \leq L \quad (6)$$

$$h_i(n) = c_x(i-L|n)/c_x(0|n), \quad i = 0, 1, \dots, 2L \quad (7)$$

where  $L$  is the maximum lag,  $0 < \alpha_1 < 1$  is so-called smoothing factor and  $p_x(n)$  is the power of  $x(n)$ , which is recursively estimated using

$$p_x(n) = \alpha_2 p_x(n-1) + (1-\alpha_2)x^2(n) \quad (8)$$

where  $0 < \alpha_2 < 1$  is another smoothing factor. In the case of Gaussian noise, the steady state of (7) is given by [4], [2]

$$h_i(\infty) = \sum_{m=1}^P \tilde{A}_m \cos(2\pi f_m(L-i)), \quad i = 0, 1, \dots, 2L \quad (9)$$

where  $\tilde{A}_m$  are positive and unknown constants. Therefore, the adaptive filter in the steady state is a narrow bandpass FIR filter whose center frequencies are equal to the frequencies of the input sinusoidal signal.

## 3. THE UNITY-GAIN CBALE AND ITS PERFORMANCE IN COLORED NON-GAUSSIAN NOISE CASE

### 3.1 The Unity-Gain CBALE

Figure 2 shows the novel UGCBALE. It is apparent that it is composed of the non-unity-gain CBALE followed by an adaptive noise canceller (ANC). The basic idea arises from the fact that the output of the non-unity-gain CBALE is given by

$$z(n) = \zeta(z)s(n) \quad (10)$$

where  $\zeta(z)$  is an unknown gain frequency dependent and the noisy signal  $x(n)$  is given by

$$x(n) = s(n) + v(n) \quad (11)$$

Therefore, both  $z(n)$  and  $x(n)$  can be taken respectively as the reference and the primary inputs of an ANC. The ANC deletes the correlated signal of both reference and primary from its output  $e(n)$ . To achieve this task, the adaptive filter associated with the ANC will provide a gain  $1/\zeta(z)$ . This in turn implies that the output of this adaptive filter will be an estimate of the sinusoidal signal  $s(n)$ , i.e.,  $y(n) = \hat{s}(n)$  and the output error of the ANC will be an estimate of the noise  $v(n)$ .

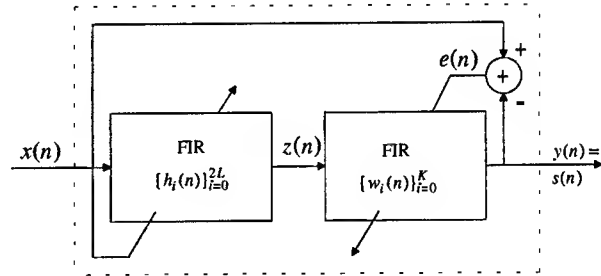


Figure 2. The unity-gain cumulant-based adaptive line enhancer (UGCBALE).

The adaptive filter associated with the ANC can be implemented as an FIR filter updated using the NLMS algorithm. That is, the output of the UGCBALE is computed by

$$y(n) = \sum_{i=0}^K w_i(n)z(n-i) \quad (12)$$

where  $z(n)$  is the output of the non-unity-gain CBALE given by (5). The adaptive coefficients  $w_i(n)$  can be updated using the NLMS written as

$$w_i(n+1) = w_i(n) + \mu e(n)z(n-i)/\gamma \quad (13)$$

where  $\mu$  is a positive step size,  $\gamma = \sum_{i=0}^K z^2(n-i)$  and  $e(n)$  is the output error of the ANC given by

$$e(n) = x(n) - y(n) \quad (14)$$

### 3.2 The Performance in Colored Non-Gaussian Noise Case

It is obvious that the suggested UGCBale is composed of two cascade connected FIR adaptive filters; the first FIR filter is updated based on higher-order statistics while the second FIR filter is updated based on second-order statistics. Therefore, the type of noise (Gaussian or non-Gaussian) affects the first filter alone. This implies that to investigate the performance of the UGCBale in non-Gaussian noise case, one needs to examine the performance of the first adaptive filter. To achieve that, assume  $v(n)$  to be colored non-Gaussian noise, then it can be shown that the steady state of (7) is given by

$$h(i) = \sum_{m=1}^P \tilde{A}_m \cos(2\pi f_m(L-i)) + \beta c_{4v}(L-i), \quad i=0,1,\dots,2L \quad (15)$$

where  $\beta$  is the reciprocal of the Kurtosis of  $x(n)$  and  $c_{4v}(\cdot)$  is a one-dimensional slice of the fourth-order cumulant of the noise  $v(n)$ . That is, the steady state impulse response in this case is sinusoidal signal corrupted by the fourth-order cumulant of the noise normalized to the Kurtosis of  $x(n)$ .

In [7], it has been proved that a signal composed of sinusoids plus noise possesses two SNR's. The first is the conventional (original) SNR associated with second-order statistics of the signal and the second is a new SNR associated with the employed one-dimensional slice of the fourth-order cumulant of the signal. From (15) and after some manipulations, the new local SNR defined as the ratio of the  $m$ th sinusoidal amplitude  $\tilde{A}_m$  to the value  $\beta c_{4v}(0)$  and termed the signal-to-noise Kurtosis ratio ( $SNKR_m$ ), is given by [6], [7]

$$SNKR_m = (12/8) |\sigma^4 / \gamma| \lambda SNR_m^2 \quad (16)$$

where the ratio  $|\sigma^4 / \gamma|$ , characterizing the white noise generating the additive colored noise, is a constant for each particular non-Gaussian noise, it is, for example, 5/6 for uniformly distributed random process (UDRP) noise [7];  $\lambda$  is a new measure for the noise spectrum distribution, it is equal to one when the noise is white and increases with the decrease of the noise spectrum bandwidth [7] and  $SNR_m$  is the conventional local SNR. This implies that in the case of UDRP noise and  $SNR_m = 2$ ,  $SNKR_m$  is equal to  $5\lambda$ . Then, if the noise is white,  $SNKR_m$  is 2.5 times  $SNR_m$ . If the noise is colored and  $\lambda = 10$ , for example,  $SNKR_m$  is 25 times  $SNR_m$ . This proves that updating the first filter using higher-order cumulants eliminates completely the effect of Gaussian noise and reduces the effect of colored non-Gaussian noise.

In white non-Gaussian noise case, the UGCBale outperforms the conventional ALE provided that the  $SNKR_m$  is equal to or greater than  $SNR_m$ . In this case, it is obvious from (16) that the conventional  $SNR_m$  is given by

$$SNR_m \geq (8/12) |\gamma / \sigma^4| \quad (17)$$

This implies that there is a minimum  $SNR_m$  characterizing each particular white non-Gaussian noise, which ensures that the UGCBale outperforms the conventional ALE. This minimum  $SNR_m$  is equal to 0.8 for uniformly distributed noise.

## 4. SIMULATION RESULTS

To examine the performance of the UGCBale in comparison with the conventional ALE, the following simulation examples are conducted. In these examples, the results are averaged over 20 trials each consists of 2048 iterations.

**Example 1-** In this example the input signal  $x(n)$  is given by

$$x(n) = \cos(0.2\pi n) + v(n) \quad (18)$$

where  $v(n)$  is a zero-mean colored uniform distributed random process (UDRP) noise generated by passing a zero-mean white UDRP noise through the following coloring filter:

$$G(z) = 0.138 \frac{1 + 2z^{-1} + z^{-2}}{(1 - 0.98e^{j0.5\pi}z^{-1})(1 - 0.98e^{-j0.5\pi}z^{-1})} \quad (19)$$

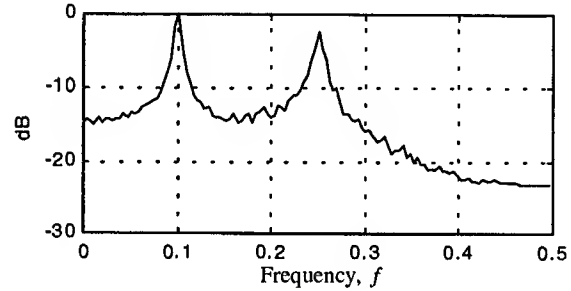


Fig. 3. Spectrum of the input signal  $x(n)$  used for Example 1.

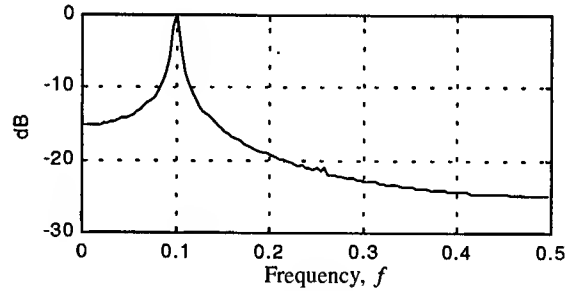


Figure 4. Spectrum of the output of the UGCBale.

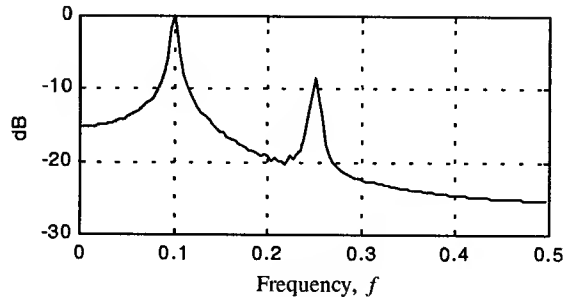


Figure 5. Spectrum of the output of the ALE.



That is, the spectrum of the colored noise has two strong peaks at frequencies  $f = \pm 0.25$ . The autocorrelation function of this noise is an exponentially damped sinusoid of damping factor 0.98 and an oscillation frequency of 0.25, which implies that the noise is highly correlated. The variance of  $v(n)$  is adjusted to achieve 2.0 SNR. Using [6], [7] the factor  $\lambda$  given in (16) of the noise spectrum provided by the coloring filter (19) can be computed. It is equal to 25.5, which means that the SNKR is about 64 times the original SNR. This SNKR gives primary indication that the UGCBale outperforms the ALE by about 18.0 dB. The parameters characterizing the UGCBale  $\lambda, \alpha, L, \mu, K$  are taken 0.99, 0.999, 24, 0.001 and 23 respectively. The parameters characterizing the conventional ALE  $M, \Delta, \mu$  are taken 36, 36, 0.001 respectively. The step sizes are selected experimentally to obtain maximum noise attenuation and stable adaptation process. Both enhancers are started with zero initial weights. Figures 3, 4 and 5 show the spectra of the input, the UGCBale output and the ALE output respectively. From these spectra, the UGCBale attenuates the noise spectral peak at  $f = 0.25$  by about 18.0 dB while the ALE attenuates this peak by about 8.0 dB. This implies that the UGCBale outperforms the ALE by about 20.0 dB, 10.0 dB for each noise spectral peak.

**Example 2-** In this example the input signal  $x(n)$  is given by

$$x(n) = \cos(0.2\pi n) + \cos(0.6\pi n) + v(n) \quad (19)$$

where  $v(n)$  is the noise described in Example 1. The local SNR's are  $SNR_1 = SNR_2 = 2.0$ . The parameters characterizing both enhancers are taken as in Example 1. Figures 6, 7 and 8 show the spectra of the input, the outputs of both the UGCBale and the ALE, respectively. It is apparent that the UGCBale still outperforms the ALE by about 16.0 dB. The impulse responses of both the ANC filter associated with the UGCBale and the ALE are investigated. The predictor filter of the ALE has two jobs, the first is to remove the noise and the second is to keep the gain equal to one at the frequencies of the sinusoidal signal. The adaptive filter of the ANC associated with the UGCBale has only one job. It tries to adjust the gain equal to one at the frequencies of the sinusoidal signal. This is because the noise was removed by the CBale. This single job facilitates the adaptation process of this adaptive filter. For space limitation, the impulse responses of all adaptive filters associated with both enhancers and error signals are omitted. Spectral estimation of each signal is obtained by using 256-FFT to 256 points of the signal before the end of simulation (before the final iteration).

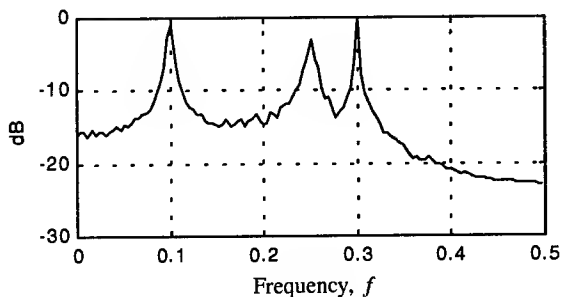


Figure 6. Spectrum of the input signal for Example 2.

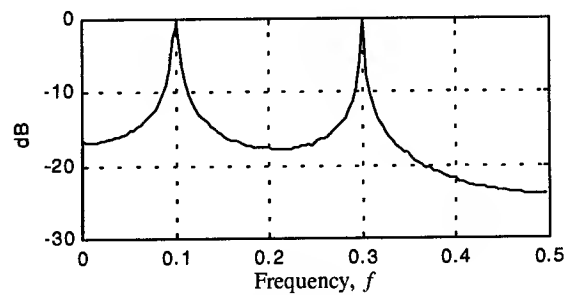


Figure 7. Spectrum of the output of the UGCBale.

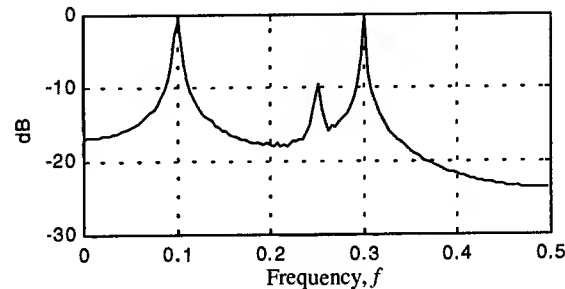


Figure 8. Spectrum of the output of the ALE.

## 5. CONCLUSION

In this paper, a unity-gain cumulant-based adaptive line enhancer (UGCBale) has been presented. It is composed of two cascade connected FIR adaptive filters. The first one is updated using higher-order cumulants of the input signal. It is then insensitive to Gaussian noise (white or colored) and it has been shown theoretically and experimentally that it performs well in colored non-Gaussian noise case. The second one, updated using second-order statistics based NLMS algorithm, is to adjust the overall gain to be approximately one. Then, adaptive sinusoidal interference canceller is available. Simulation results have shown that the UGCBale outperforms the conventional ALE having the same number of coefficients in the case of colored uniformly distributed random process (UDRP) noise.

## 6. REFERENCES

- [1] B. Widrow *et al.*, "Adaptive noise canceling: Principles and applications," *Proc. IEEE*, vol. 63, pp. 1692-1716, Dec. 1975.
- [2] H. M. Ibrahim, R. R. Gharieb and M. M. Hassan, "A higher order statistics based adaptive algorithm for line enhancement," *IEEE Trans. Signal Processing*, vol. 47, pp. 527-531, Feb. 1999.
- [3] H. M. Ibrahim and R. R. Gharieb, "Two-dimensional cumulant-based adaptive enhancer," *IEEE Trans. Signal Processing*, vol. 47, pp. 593-596, Feb. 1999.
- [4] R. R. Gharieb, "Development and applications of adaptive filtering techniques using higher-order statistics," Ph.D. thesis, Assiut University, Assiut, Egypt, April 1997.
- [5] R. R. Gharieb, "Cumulant-based LP method for two-dimensional spectral estimation," *IEE Proc. Vision, image and Signal Processing*, vol. 146, pp. 307-312, Dec. 1999.

- [6] R. R. Gharieb, "New results on employing cumulants for retrieving sinusoids in colored non-Gaussian noise," *IEEE Trans. Signal Processing* (to appear July 2000)
- [7] R. R. Gharieb, "Higher order statistics based IIR notch filtering scheme for enhancing sinusoids in colored noise," *IEE Proc. Vision, Image and Signal Processing*. (To appear)
- [8] R. R. Gharieb, Y. Horita and T. Murai, "Retrieving sinusoids in colored Raleigh noise by a cumulant-based FBLP approach," *Proc. ICASSP'2000, Istanbul, Turkey*, pp. 741-743.
- [9] B. Widrow and S. Stearns, *Adaptive signal processing*. Englewood Cliffs, NJ: Prentice-Hall, 1985.
- [10] C. L. Nikias and A. P. Petropulu, *Higher order spectral analysis: A nonlinear signal processing framework*. Englewood Cliffs, NJ: Prentice-Hall, 1993.
- [11] J. M. M. Anderson, G. B. Ginakakis and A. Swami, "Harmonic retrieval using higher-order statistics: A deterministic formulation," *IEEE Trans. Signal Processing*, vol. 43, pp. 2729-2741, 1994.

# ADAPTIVE DETECTION AND EXTRACTION OF SPARSE SIGNALS EMBEDDED IN COLORED GAUSSIAN NOISE USING HIGHER ORDER STATISTICS

R. R. Gharieb<sup>†1</sup>, A. Cichocki<sup>1,2</sup> and S. F. Filipowicz<sup>2</sup>

1 Lab. for Advanced Brain Signal Processing

Brain Science Institute, RIKEN

2-1 Hirosawa, Wako-Shi, Saitama 351-0198, JAPAN

E-mail: [reda@bsp.brain.riken.go.jp](mailto:reda@bsp.brain.riken.go.jp), Fax: + 81-48-467-9694

2 Warsaw University of Technology, Warsaw, Koszykowa 75, IETME, POLAND

## ABSTRACT

A cumulant-based adaptive approach for the detection and extraction of sparse signal embedded in colored Gaussian noise is presented. In this approach, the extracted signal is obtained by adaptive FIR filtering of the noisy signal. Coefficients of the adaptive filter are updated using a recursive algorithm based on a sum of cumulants of orders  $k \geq 3$  of the input signal. This is to ensure super sufficient detection of different sparse signals and to ensure efficient removal of colored Gaussian noise. It is shown that when the sparse pulse is absent, the coefficients of the adaptive filter converge to zero. However, when the sparse pulse exists the FIR adaptive filter converges to a type of signal-matched filters. Simulation and experimental results are included to show the high efficiency of the presented approach in comparison with the adaptive short-term correlation counterpart.

**Keywords-** Higher-order Statistics, Sparse Signals, Biomedical signals, Gaussian noise, Detection.

## 1. INTRODUCTION

Extraction of short-time extent (sparse) signal embedded in additive noise is a problem frequently encountered in a variety of fields such as biomedical signal processing, radar, communications, etc. Various adaptive filters based on the correlation of the input noisy signal may be satisfactory when the additive noise is white. However, in highly correlated noise case, the impulse response of the adaptive filter converges to the autocorrelation function of both the signal plus the additive colored noise. This implies that the passband of the filter spectrum will be in the same band of both the signal and noise. Therefore, the adaptive filter output will be a version of input noisy signal [1], [2].

Recently, higher-order statistics or cumulants have been successfully employed for the detection and classification of non-Gaussian signals in Gaussian noise. This is because higher-order cumulants of Gaussian noise either white or colored are identically zero [3]-[12]. Various fourth-order cumulant slices of the noisy signal have been used for the retrieval of harmonic signal in colored Gaussian noise. It has also been explained that employing fourth-order cumulants slices is an efficient approach to handling colored non-Gaussian noise corrupting a sinusoidal signal [4], [8]. Various fourth-order cumulant-based filtering techniques (fixed or

adaptive) have been described for the enhancement of a sinusoidal signal in colored either Gaussian or non-Gaussian noise. These techniques have been developed with the assumption that the signal is stationary [3]-[5]. In [3], coefficients of an FIR adaptive line enhancer have been recursively updated using one-dimensional slice of the fourth-order cumulant of the input signal. Computation of this slice needs the power of the input signal to be recursively estimated. Therefore any small error of the estimated power of the input signal will influence the performance of the algorithm, especially when the signal is nonstationary. In [5], a cumulant-based IIR adaptive notch filter has been described for the enhancement and tracking of a single sinusoid in noise.

In this paper, a new approach for the detection and extraction of sparse signals embedded in colored Gaussian noise is presented. In this approach, the extracted signal is obtained by passing the noisy signal through an FIR adaptive filter whose coefficients are updated using a proposed algorithm based on a sum of cumulants of orders  $k \geq 3$  of the input signal. This is to ensure reliable and efficient detection of sparse signal and to ensure the removal of Gaussian noise. Another important motivation of employing weighted sum of cumulants is to avoid a problem may arise when specific higher-order cumulant is zero.

## 2. SIGNAL MODEL

In this paper we concerned with a class of signals that can be modeled as a sum of short-extent pulses (sparse signals), i.e.,

$$s(n) = \sum_{i=0}^I A_i \delta(n - M_i) \exp(-\alpha_i n^2) \cos(\omega_i n + \varphi_i) \quad (1)$$

where  $A_i, \alpha_i, \omega_i$  and  $\varphi_i$  are unknown parameters: amplitude, the damping factor, the frequency and the phase of the  $i$ th cosine pulse, respectively, with  $0 \leq \omega_i \leq \pi$ . The time  $M_i$  represents the time position of the center of the  $i$ th pulse on the time axis  $n$  and  $\delta(\cdot)$  is the delta function.

Due to the presence of noise, one observes a contaminated version of  $s(n)$ , namely

$$x(n) = s(n) + v(n) \quad (2)$$

where  $v(n)$  is assumed to be a zero-mean additive Gaussian noise of unknown covariance. Additionally,  $v(n)$  is considered to be the output of a stable, linear time-invariant (LTI) filter

<sup>†</sup> Dr. R. R. Gharieb is on leave from the Faculty of Engineering, Assiut University, Egypt

driven by i.i.d. Gaussian noise with bounded higher-order moments. A local signal-to-noise ratio (SNR) is defined as the ratio of the maximum amplitude of the pulse signal to noise power, i.e.,

$$SNR_i = 10 \log_{10} (I A_i / \sigma_v^2) \quad (3)$$

The objective is to detect and to extract  $s(n)$  given only the noisy signal  $x(n)$ .

### 3. CUMULANT-BASED FIR ADAPTIVE FILTERING

#### 3.1 FIR Fixed Filtering

Our idea is to pass the noisy signal through an FIR filter whose impulse response is a sum of higher-order cumulants of  $x(n)$ . The noise-free signal is deterministic signal consisting of a sum of pluses (sparse signal). This sparse signal is embedded in colored Gaussian noise. In such case, mixed cumulants [10] computed over time average have been adopted. The second- and third order-cumulants are given by [9]

$$\bar{c}_{2x}(\tau) = \langle u(n)u(n+\tau) \rangle = \frac{1}{N} \sum_{n=0}^{N-1} u(n)u(n+\tau) \quad (4)$$

$$\begin{aligned} \bar{c}_{3x}(\tau_1, \tau_2) &= \langle u(n)u(n+\tau_1)u(n+\tau_2) \rangle \\ &= \frac{1}{N} \sum_{n=0}^{N-1} u(n)u(n+\tau_1)u(n+\tau_2) \end{aligned} \quad (5)$$

where  $u(n) = x(n) - \langle x(n) \rangle$ .

For convenience, let in (1)  $I$ ,  $M_i$ , and  $\phi_i$  be equal to zero, i.e., the noise-free signal is only one pulse at the origin time with zero phase. In this simple case we use the third-order mixed cumulant of  $x(n)$ . In this case, the impulse response of the FIR adaptive filter is suggested to be:

$$h(\tau) = \bar{c}_{2x}(0, \tau) = \bar{c}_{3s}(0, \tau) + c_{3v}(0, \tau) \quad (6)$$

where  $c_{3v}(\cdot)$  is the ensemble third-order cumulant of the noise  $v(n)$ . Due to the fact that the noise is Gaussian,  $h(\tau)$  in (6) reduces to

$$h(\tau) = \bar{c}_{3s}(0, \tau) \quad (7)$$

Using (1), (5) and (7),  $h(\tau)$  can be written as

$$\begin{aligned} h(\tau) &= \frac{1}{N} \sum_{n=0}^{N-1} A^4 e^{-2\alpha n^2} e^{-\alpha(n+\tau)^2} \\ &\quad \times (\cos(\omega_0 n))^2 \cos(\omega_0(n+\tau)) \end{aligned} \quad (8)$$

After simple manipulations the impulse response can be given by

$$h(\tau) = \gamma A^4 e^{-\alpha \tau^2} \cos(\omega_0 \tau) \quad (9)$$

where  $\gamma$  is a constant fixed or slowly changing with time shift  $\tau$ . Then when the pulse exits, the impulse response of the FIR filter is a type of pulse-matched filters, which means that the filter bandpass is identical with the band of the pulse signal. If the pulse is absent, the impulse response is identically zero, which implies that in this case the output is zero.

#### 3.2 FIR Adaptive Filtering

Because fixed filtering is not satisfactory to deal with nonstationary signals, adaptive filter is essentially required. Therefore, for tracking ability, it is desired to formulate an FIR adaptive filter based on the idea described in subsection 3.1. This adaptive filter is required to detect the existence and/or the absence of each pulse. In the existence case, it is required that the adaptive filter impulse response is to converge to the pulse-matched function. And in the absence case, it is required that the adaptive filter is to forget the values of the impulse response previously computed. We propose that the impulse response is to be recursively computed as follows

$$\begin{aligned} h(\tau | n) &= \beta h(\tau | n) + (1 - \beta) F[x^2(n)] \\ &\quad \times x(n - P + \tau), \quad \tau = 0, 1, \dots, 2P \end{aligned} \quad (10)$$

where  $0 < \beta < 1$  is so-called forgetting factor,  $F[x^2(n)]$  is a nonlinear function of  $x^2(n)$  and  $x(n)$  is the input signal with mean removed. Due to this nonlinear function, the impulse response of the adaptive filter converges to a sum of cumulants of orders  $k \geq 3$ .

The output of the adaptive filter is given by

$$y(n) = \sum_{\tau=0}^{2P} \text{Sign}(h(P | n)) h(\tau | n) x(n - P - \tau) \quad (11)$$

where  $\text{Sign}(h(P | n))$  is the sign of  $h(P | n)$  given in (10). This sign is included to avoid the negative sign may appear with higher order cumulants (i.e., Skewness, Kurtosis, etc.).

Figure 1 shows a block diagram for the cumulant-based adaptive filter while Figure 2 shows an illustrative implementation of the cumulant-based recursive algorithm given in (10). It is obvious that the nonlinear function make us be able to use the adaptive short-term correlation estimator for the computation of a sum of cumulants of orders  $k \geq 3$  of the input signal [2].

It is worth to note that the rate of the recursive algorithm is dependent upon the choice of the factor  $\beta$ . Small values cause fast forgetting but on the other hand it may cause insufficient smoothing, i.e., not enough convergence to the pulse signal shape. Therefore, selecting  $\beta$  is dependent upon trail work and upon the spread of the signal pulses.

For convenience, the counterpart of the presented approach, which is based on second-order statistics and termed the ASC algorithm, can be summarized as follows [2]

$$h(\tau | n) = \beta h(\tau | n) + (1 - \beta) x(n) x(n - P + \tau), \quad \tau = 0, 1, \dots, 2P \quad (12)$$

This implies that the impulse response of the adaptive filter converges to

$$h(\tau) = \bar{r}_{2s}(\tau) + r_{2v}(\tau) \quad (13)$$

Because we assume that the noise is colored (especially highly colored), the impulse response is equal to the autocorrelation function of the noise when the pulse signal is absent and is equal to the autocorrelation of the pulse signal plus the autocorrelation of the noise when the pulse signal exists. This implies that the ASC algorithm will not be able to reject the colored noise even whenever the signal pulse is absent or existing.

## 4. SIMULATION RESULTS

To examine the presented adaptive filtering techniques for the enhancement and detection of sparse signal in colored Gaussian noise, the following examples have been conducted. Results of the presented algorithm are compared with the adaptive short-term correlation algorithm. In all examples the additive colored Gaussian noise  $v(n)$  is generated by passing zero-mean, i.i.d. Gaussian noise through the following coloring filter:

$$0.1007 \frac{1 + 2z^{-1} + z^{-2}}{(1 - 0.98e^{j2\pi 0.2})(1 - 0.98e^{-j2\pi 0.2})} \quad (14)$$

The covariance of this coloring filter is a damped sinusoid with damped factor 0.98 and an oscillation normalized frequency of 0.2. This implies that the autocorrelation (second-order statistics of  $v(n)$ ) is of considerable values over long time shift  $\tau$ . The noisy signal  $x(n)$  of length 2000 is obtained by adding the colored noise  $v(n)$  to the sparse signal that is specified with every example. The order of FIR adaptive filters for both algorithms is taken  $P=8$  and the forgetting factors for both algorithms is  $\beta=0.95$ . The nonlinear function is taken

$$F(x^2) = 1/(1 + e^{-0.5x^2}).$$

**Example 1:** In this example the noise-free sparse signal (amplitude versus time) shown in Figure 3 (a) is investigated. The power of the noise  $v(n)$  is adjusted to achieve 0.0 dB SNR. The signal embedded in colored noise is shown in Figure 3 (b). Figures 3 (c) and (d) show the results of estimated signals using both algorithms obtained from 20 Monte Carlo runs. It is obvious that the presented algorithm performs better than the one based on correlations.

**Example 2:** In this example the noise-free sparse rectangular signal shown in Figure 4 (a) is investigated. The power of the noise  $v(n)$  is adjusted to achieve 0.0dB SNR. The sparse signal embedded in noise is shown in Figure 4 (b). Figures 4 (c) and (d) show the results of enhancement using both algorithms obtained from 20 Monte Carlo runs. It is obvious that the presented algorithm still performs better than the one based on correlations.

**Example 3:** In this example we have used ECG (electrocardiogram artifact) recorded by MEG machine. Only one channel is shown in Fig. 5 (a). Figures 5 (b) and (c) show the results of both algorithms obtained from 20 channels. It is apparent that the presented algorithm outperforms the one based on correlations.

## 6. CONCLUSION

A cumulant-based adaptive approach for the detection and extraction of sparse signal embedded in colored Gaussian noise has been presented. In this approach, the noisy signal is passed through an FIR adaptive matched filter whose coefficients are updated using a recursive algorithm based on a sum of cumulants of orders  $k \geq 3$  of the input signal. This is to ensure super sufficient classification of various signals and to ensure the removal of Gaussian noise. It has been shown that in the absent of the sparse signal, the coefficients of the adaptive filter converge to zero. However, the adaptive filter converges to a type of sparse-matched filters over the sparse time window. Simulation and experimental results have shown the efficiency of the presented approach in comparison with the adaptive short-term correlation counterpart.

## 5. REFERENCES

- [1] B. Widrow *et al.*, "Adaptive noise canceling: Principles and applications," *Proc. IEEE*, vol. 63, pp. 1692-1716, Dec. 1975.
- [2] N. Ahmed and S. Vijayendra, "An algorithm for line enhancement," *Proc. IEEE*, vol. 70, pp. 1459-1460, 1982.
- [3] H. M. Ibrahim, R. R. Gharieb and M. M. Hassan, "A higher order statistics based adaptive algorithm for line enhancement," *IEEE Trans. Signal Processing*, vol. 47, pp. 527-531, Feb. 1999.
- [4] R. R. Gharieb, "Higher order statistics based IIR notch filtering scheme for enhancing sinusoids in colored noise," *IEE Proceedings-Vision, Image and Signal processing*. (To appear)
- [5] R. R. Gharieb, Y. Horita and T. Murai, "Cumulant-based IIR adaptive notch filter for enhancing and tracking a single sinusoid in noise," Submitted to *IEE Proceedings-Vision, Image and Signal Processing*.
- [6] R. R. Gharieb, "Cumulant-based LP method for two-dimensional spectral estimation," *IEE proceedings-Vision, Image and Signal Processing*, vol. 146, pp. 307-312, Dec. 1999.
- [7] R. R. Gharieb, "New results on employing cumulants for retrieving sinusoids in colored non-Gaussian noise," *IEEE Trans. Signal Processing* (To appear July 2000).
- [8] J. M. Mendel, "Tutorial on higher-order statistics (spectra): Theoretical results and some applications," *Proc. IEEE*, vol. 79, pp. 278-305, 1991.
- [9] C. L. Nikias and A. P. Petropulu, *Higher order spectral analysis: A nonlinear signal processing framework*. Englewood Cliffs, NJ: Prentice-Hall, 1993.
- [10] J. M. M. Anderson, G. B. Ginnakis and A. Swami, "Harmonic retrieval using higher order statistics: A deterministic formulation," *IEEE Trans. Signal Processing*, vol. 43, pp. 2729-2741, 1994.
- [11] B. M. Sadler, G. B. Ginnakis and K. Lii, "Estimation and detection in non-Gaussian noise using higher-order statistics," *IEEE Trans. Signal Processing*, vol. 42, pp. 2729-2740, Oct. 1994.
- [12] G. B. Ginnakis and M. K. Tsatsanis, "Signal detection and classification using matched filter and higher-order statistics," *IEEE Trans. Signal Processing*, vol. 38, pp. 1284-1296, July 1990.

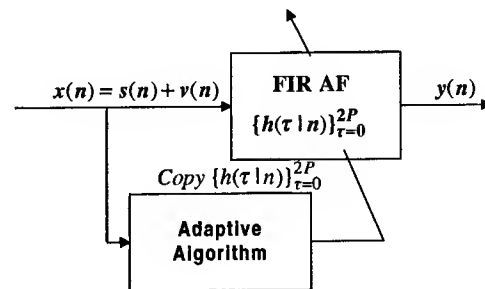


Figure 1. Block diagram of adaptive filtering based on the suggested cumulants based adaptive algorithm.

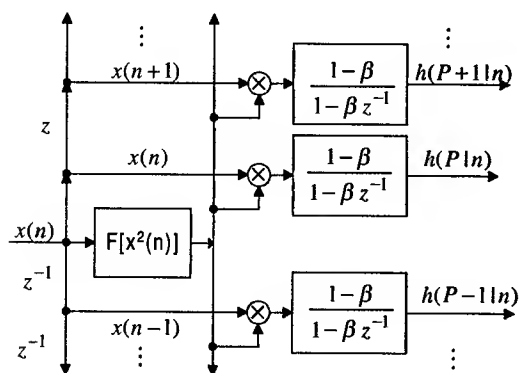


Figure 2. A scheme using nonlinear function and a bank of smoothing filters for the recursive computation of the adaptive filter coefficients using a sum of cumulants of orders  $k \geq 3$  of the input signal.

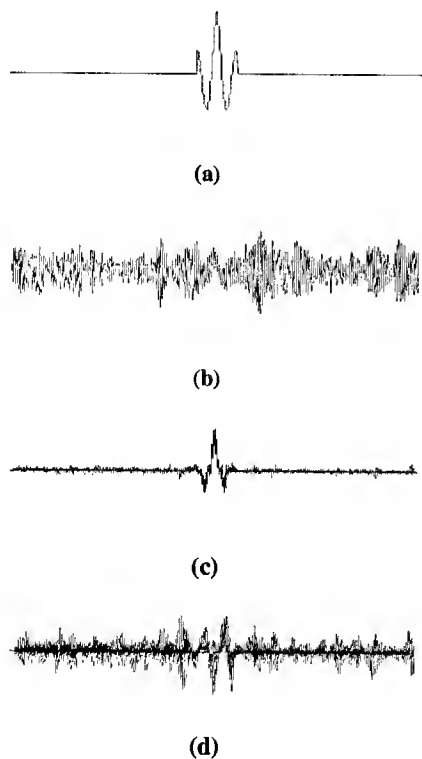


Figure 3. Results of Example 1: (a), noise-free sparse signal; (b), observed signal with additive noise; (c), reconstructed signal using the proposed technique; and (d), reconstructed signal using the conventional ASC algorithm.

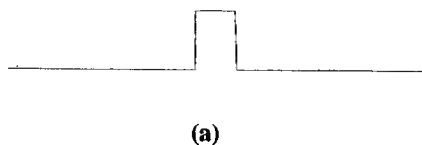


Figure 4. Results of Example 2: (a), noise-free rectangular signal; (b), observed noisy signal; (c), the enhanced output using the proposed technique; and (d), the output of the conventional ASC algorithm.

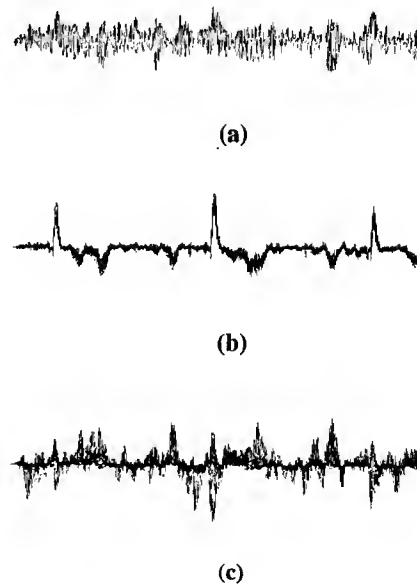


Figure 5. Results of Example 3 (ECG): (a), recorded ECG signal by using MEG machine; (b), reconstructed signal using the proposed technique; (c), reconstructed signal using the conventional ASC algorithm.

# Higher-Order Matched Field Processing

Reza M. Dizaji, R. Lynn Kirlin, and N. Ross Chapman\*

Electrical and Computer Engineering Department, \*School of Earth and Ocean Sciences, University of Victoria, Victoria, BC, V8W 3P6, Canada email: kirlin@ece.uvic.ca

## Abstract

This paper introduces a cross-relation (CR) based higher order matched field (MF) processing technique for estimating the location of a random source in shallow water. It is known that the probability density function (PDF) signals emitted from marine vessels have higher order components. Using the higher order MF processor can cancel the effect of either white or non-white Gaussian random interferences since the third and higher odd moments and third and higher order cumulants of Gaussian random interferences are zero. We have examined the higher order content of experimental ship data as it effects the MF processor for estimation of location using different frequency bands. The ship data is recorded during a sea trial conducted on September 1993 in a region close to Vancouver Island, BC, Canada.

## 1. Introduction

Parametric estimators based on matching between the measured signals and replicas based on the environmental model are widely used in various signal processing applications including source location and environmental air channel parameter estimation [1,2]. The development of these techniques, well known as matched field processing (MFP) is indebted to reliable numerical models that model the field with high precision. In underwater acoustics, several software packages have been released, such as SAFARI, OASES and ORCA [3,4] for modeling the acoustic field.

The higher-order statistics have shown wide applicability in many diverse fields such as sonar, radar, seismic signal processing, data analysis and system identification [5-7]. Specifically, cumulants and their associated Fourier transforms, known as polyspectra, reveal not only amplitude information but also phase information. This is important because, as is well known, second-order statistics (such as the auto-correlation) are phase blind. Cumulants, on the other hand, are blind to any kind of Gaussian process; thus they can handle colored Gaussian measurement noise automatically, whereas correlation-based methods do not. Cumulant-based methods boost signal-to-noise ratio when signals are corrupted by Gaussian interference. Ship data has a complex distribution with statistics higher than second order, so a matched field processor based on higher order statistics will let us use more of the information in the data. The greatest drawbacks to the use of higher-order statistics are that they require longer data records and much more computation than do correlation-based methods. Longer data lengths are needed in order to reduce the variance associated with estimating the higher-order statistics from real data using sample averaging techniques.

The  $k$ th-order cumulant is defined [8] in terms of its joint moments of orders up to  $k$  and vice versa. The moment-to-cumulant formula is

$$C_x(I) = \sum_{\bigcup_{p=1}^q I_p = I} (-1)^{q-1} (q-1)! \prod_{p=1}^q m_x(I_p) \quad (1)$$

where  $\bigcup_{p=1}^q I_p = I$  denotes summation over all partitions of set  $I$ . Set  $I$  contains the indices of the components of vector  $x$  where  $x = [x_1, x_2, \dots, x_k]^T$  denotes a collection of random variables. The partition of the set  $I$  is the unordered collection of nonintersecting nonempty sets  $I_p$  such that  $\bigcup_{p=1}^q I_p = I$  where  $q$  is the number of partitions sets  $I_p$ .  $m_x(I_p)$  indicates the moment of the partition  $x$  corresponding to set  $I_p$ , i.e.,  $m_x(I_p) = E(x_1 x_2 \dots x_p)$ .

The cumulant-to-moment formula is:

$$m_x(I) = \sum_{\bigcup_{p=1}^q I_p = I} C_x(I_p) \quad (2)$$

## 2. Higher-order MFP

Let us consider the geometry of the measurement system for ship localization in shallow water using a vertical linear array with  $N$  sensors as shown in Fig. 1. This system can be modeled by a multi-channel system shown in Fig. 2, consisting of  $N$  linear transfer functions. The transfer function  $h_i$  corresponds to the paths traveled by acoustic waves from the ship to the  $i$ th sensor, including interactions with ocean bottom and surface. It is assumed that the noise is additive and is spatially and temporally white, Gaussian and uncorrelated with the input signal. The cross-relation in equation (3) follows from the linearity of the transfer functions:

$$\begin{cases} y_p(n) = h_p(n; \alpha) * S(n) \\ y_q(n) = h_q(n; \alpha) * S(n) \end{cases} \Rightarrow h_p(n; \alpha) * y_q(n) = h_q(n; \alpha) * y_p(n)$$

$$p, q = 1, 2, \dots, N; p \neq q$$

(3)

Now, we derive a matched field processor based on higher-order statistics by multiplying both sides of the DFT of equation (3) (in frequency variable  $F$ ) by a subset of  $Y_k(F)$ ,  $k = 1, \dots, N$ ,  $k \neq p, q$ . Taking expectation of

this product produces a  $T$ -order cross-relation equation ( $T \leq N$ ):

$$\begin{aligned} H_p(F; \alpha) E(Y_q(F) Y_m(F) \dots Y_{m+T-2}(F)) &= \\ H_q(F; \alpha) E(Y_p(F) Y_m(F) \dots Y_{m+T-2}(F)) \\ \Rightarrow H_p(F; \alpha) m_r(I_q) &= H_q(F; \alpha) m_r(I_p) \end{aligned} \quad (4)$$

where

$$p, q, m = 1, 2, \dots, N; p \neq q,$$

$$I_q = \{q, m, m+1, \dots, m+T-2\},$$

$$I_p = \{p, m, m+1, \dots, m+T-2\}, \text{ and MFP order is } T.$$

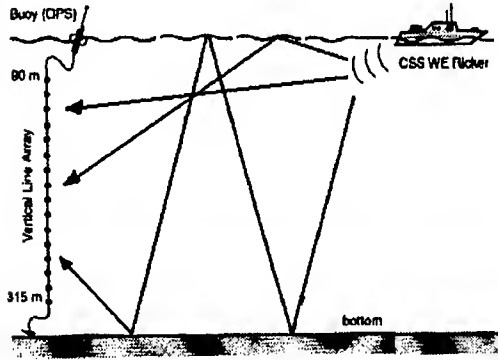


Fig. 1. The benchmark of a system for ship localization using a vertical linear array in shallow water

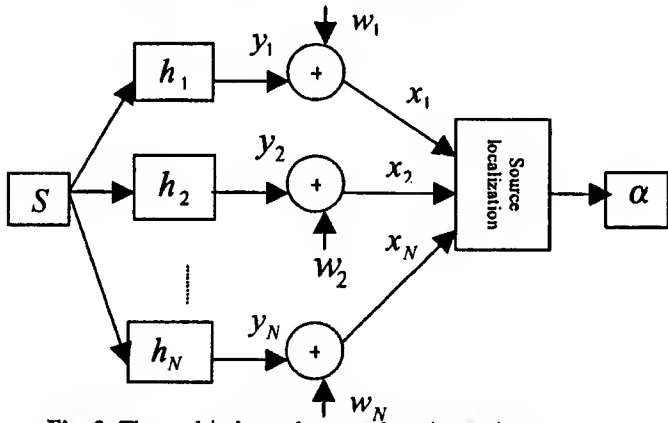


Fig. 2. The multi-channel source location estimator

If we replace moments by cumulants in equation (4) we obtain

$$H_p(F; \alpha) \left( \sum_{U_{r=1}^{I_r=I_p}} C_{Y(F)}(I_r) \right) = H_q(F; \alpha) \left( \sum_{U_{r=1}^{I_r=I_q}} C_{Y(F)}(I_r) \right) \quad (5)$$

The above equations can be written in the following matrix form to solve for all channel responses simultaneously:

$$CUM_Y H = 0 \quad (6)$$

where

$$H = [H_1^T, H_2^T, \dots, H_N^T]^T,$$

$$H_i = [H_i(0), H_i(F), \dots, H_i((L-1)F)]^T, i = 1, 2, \dots, N$$

$$CUM_Y = \begin{bmatrix} cum_{1,2,\dots,T}^T, \dots, cum_{p,q,k_1,\dots,k_{T-2}}^T, \dots \\ \left( \begin{matrix} N \\ T \end{matrix} \right) \text{ blocks} \end{bmatrix}^T$$

$$cum_{p,q,k_1,\dots,k_{T-2}}^{1 \times NL}$$

$$= \begin{bmatrix} 0 & cum_{Y_p, Y_{k_1}, \dots, Y_{k_{T-2}}} & 0 & -cum_{Y_p, Y_{k_1}, \dots, Y_{k_{T-2}}} & 0 \\ 1 \times (p-1)L & 1 \times L & 1 \times (q-p-1)L & 1 \times L & \end{bmatrix}$$

$$cum_{Y_p, Y_{k_1}, \dots, Y_{k_{T-2}}}$$

$$= \begin{bmatrix} \sum_{U_{r=1}^{I_r=I_p}} C_{Y_p(0), Y_{k_1}(0), \dots, Y_{k_{T-2}}(0)}(I_r) & \dots \\ \dots & \sum_{U_{r=1}^{I_r=I_p}} C_{Y_p((L-1)F), Y_{k_1}((L-1)F), \dots, Y_{k_{T-2}}((L-1)F)}(I_r) \end{bmatrix}$$

The identifiability condition is that the null space dimension of matrix  $CUM_Y$  should be one and  $H_i, i = 1, 2, \dots, N$ , should not be zero. To give more explicit expressions and provide more insights into the characteristics of the channels and the source signal, the following conditions are given (based on the theorem):

1. For all frequencies  $f_i, i = 1, 2, \dots, M$ , the transfer functions  $H_i, i = 1, 2, \dots, N$ , should not be zero.
2. In order to have the null space dimension of  $CUM_Y$  equal to one, and, assuming the condition mentioned above is satisfied, the source  $T$ -order moment should be non-zero for all frequencies.

For the case where channels are corrupted by noise, the least-square estimator, referred to as the high-order cross-relation based MFP, is

$$P_{Y, \alpha}^{H-CR} = \|CUM_Y H\|^{-1} \quad (7)$$

The above can be rewritten in the following form to give a more explicit expression of the processors:



$$P_{Y,\alpha}^{H-CR} = \frac{1}{\sum_p \sum_q \sum_m |H_p(F; \alpha) m_q(I_q) - H_q(F; \alpha) m_p(I_p)|^2}$$

$$= \frac{1}{\sum_p \sum_q \sum_m \left| H_p(F; \alpha) \sum_{\substack{I_r=I_p \\ I_r \neq q}} C_{Y(F)}(I_r) - H_q(F; \alpha) \sum_{\substack{I_r=I_q \\ I_r \neq p}} C_{Y(F)}(I_r) \right|^2}$$

(8)

where  $p, q, m, r$  are not equal numbers chosen from set  $\{1, 2, \dots, N\}$ .

### 3. Sensitivity analysis

Let us assume that a deviation in the true source location or environmental parameter has occurred. The cross-relation term (equation (8)) takes the form

$$CR_{pq} = E(S^T) H_m(F) \dots H_{m+T-2}(F) \mu_{pq}$$

$$\mu_{pq} = H_p(F; \alpha) H_q(F) - H_q(F; \alpha) H_p(F)$$

(9)

For parameters with low sensitivity to the pressure field, there is no considerable change in the amplitude of the transfer function. In this case we mainly focus on the transfer functions' phase. Moreover, let us assume that the array length is small enough in comparison to the water depth so with good approximation we can assume that the amplitudes of the transfer functions appearing in the formulation are the same. Equation (9) can be simplified to

$$|CR_{pq}| \approx |E(S^T(F))| |H(F)|^{T-1} |\mu_{pq}| \quad (10)$$

By substituting equation (10) in the MFP formulation (equation (8)) we have:

$$P_{Y,\alpha}^{H-CR} = \frac{1}{M(T) |H(F)|^{2(T-1)} |E(S^T(F))|^2 \sum_p \sum_q |\mu_{pq}|^2} \quad (11)$$

where  $M(T)$  is a constant multiplier equalling the number of CR terms in the MFP formulation. For an array with  $N$  sensors

$$\text{we have } M(T) = \binom{N-2}{T-1}.$$

To obtain a simpler equation, let us assume that the deviation value due to the mismatch is independent of the sensors  $p$  and  $q$ :

$$P_{Y,\alpha}^{H-CR} \approx \frac{1}{M(T) |H(F)|^{2(T-1)} |E(S^T(F))|^2 \binom{N}{2} \mu^2} \quad (12)$$

Now, let us define the MFP sensitivity. The sensitivity function  $S$  is defined as

$$S = [S_1, S_2, \dots, S_q] \quad (13)$$

where

$$S_i = \left| \frac{\partial P_{Y,\alpha}^{H-CR}}{\partial \alpha_i} \right|; i = 1, \dots, q \quad (14)$$

The sensitivity function from equation (12) becomes

$$S_i^T \approx \frac{2}{M(T) |H(F)|^{2(T-1)} |E(S^T(F))|^2 \binom{N}{2} \mu^3} \frac{\partial \mu}{\partial \alpha_i} \quad (15)$$

To see how the MFP sensitivity changes with increasing the order from  $T$  to  $T+1$  we obtain

$$S_i^{T+1} = \frac{M(T) |E(S^T(F))|^2}{M(T+1) |H(F)|^2 |E(S^{T+1}(F))|^2} S_i^T$$

$$= \frac{T |E(S^T(F))|^2}{|H(F)|^2 |E(S^{T+1}(F))|^2 (N-T-1)} S_i^T \quad (16)$$

The transfer functions norms represent the transmission loss from the source to the vertical array sensors that are relatively small because of the high ocean attenuation. This fact makes the sensitivity function to have a potentially large value for higher order MFP; however, in order to calculate the sensitivity function we need to know the relative value of the moments.

### 4. Experimental results

We have applied the 2<sup>nd</sup>, 3<sup>rd</sup>, and 4<sup>th</sup> order cross relation based MF processor to experimental ship data to examine the existence and effects of higher order content for source localization in two different frequency bands. The ship position from GPS data was at a range of 3.33km with a bearing of 153.27 degrees to the vertical linear array location. The analysis has been carried out over 73-133Hz with resolution 2Hz and over 150-270Hz with resolution of 4Hz. The replica or modeled fields used in the analysis is calculated using ORCA [3]. A towed, lower depth, acoustic beacon emits tones out of these bands, but some harmonics show in our results at the lower depth.

### 5. Conclusions

We have introduced a cross-relation (CR) based higher order matched field (MF) processing technique for estimating the location of a random source (ship) in shallow water, and in the process we have found information with regard to its higher order features which may be useful for detection or classification. It has been verified that the probability density function (PDF) of signals emitted from marine vessels have higher order components.

Use of the higher order MF processor can cancel the effect of For frequency band 73-133Hz, the 2<sup>nd</sup>, 3<sup>rd</sup> and 4<sup>th</sup> order MF processors are shown in Figs 3, 4, and 5 respectively. (Note:

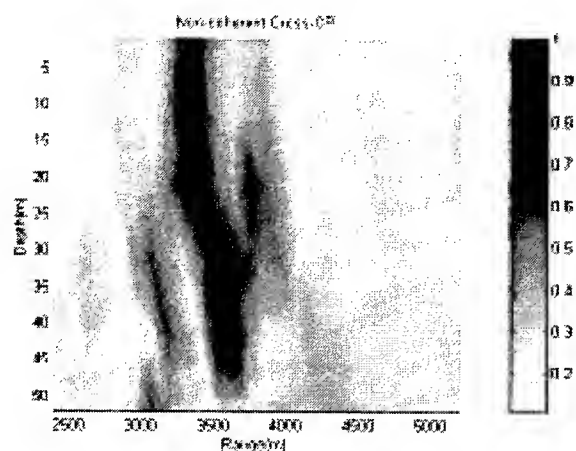


Fig. 3 Ambiguity surface for the 2<sup>nd</sup> order cross-CR processor (73-133Hz)

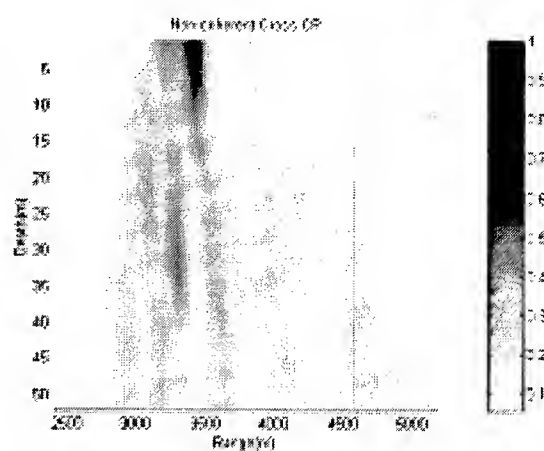


Fig. 6 Ambiguity surface for the 2<sup>nd</sup> order cross-CR processor (150-270Hz)

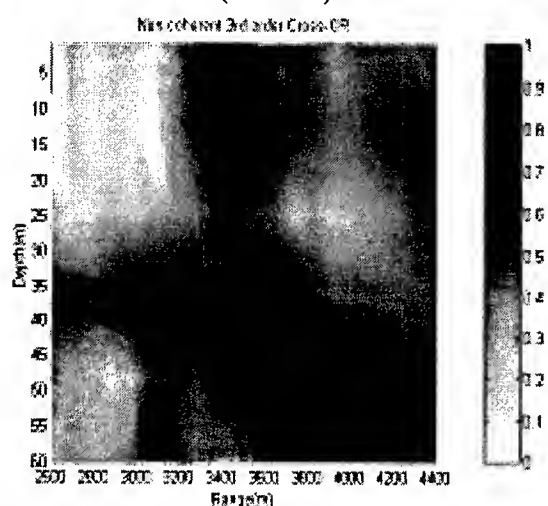


Fig. 4 Ambiguity surface for 3<sup>rd</sup> order cross-CR processor (73-133Hz)

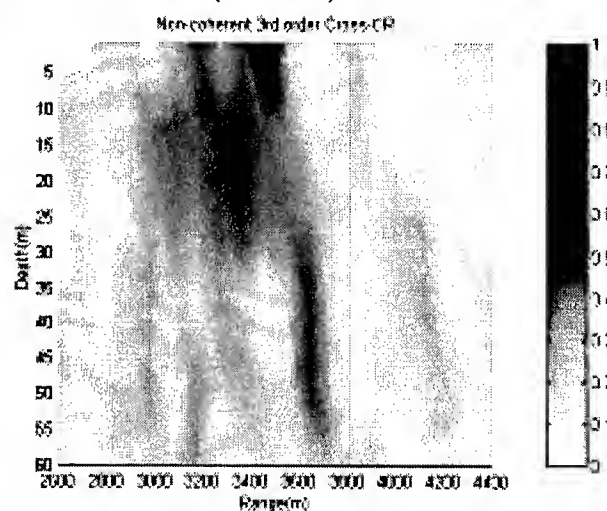


Fig. 7 Ambiguity surface for 3<sup>rd</sup> order cross-CR processor (150-270Hz)

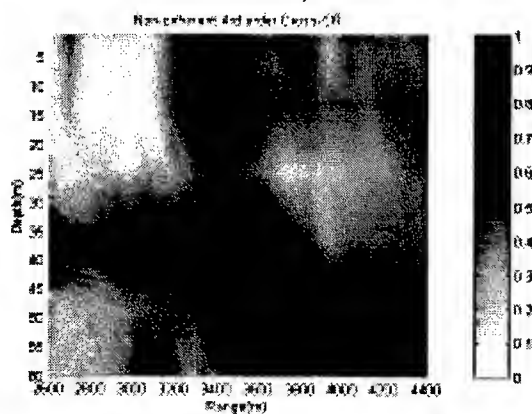


Fig. 5. Ambiguity surface for 4<sup>th</sup> order cross-CR processor (73-133Hz)

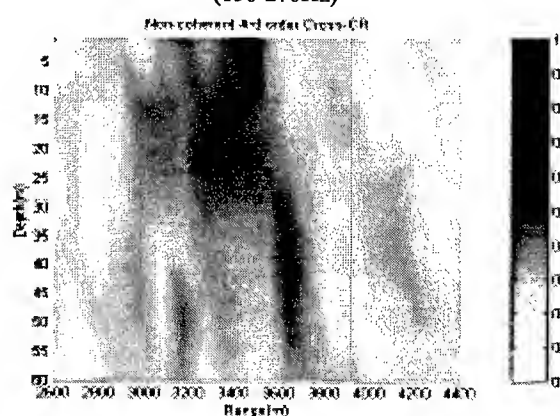


Fig. 8 Ambiguity surface for 4<sup>th</sup> order cross-CR processor (150-270Hz)

for color figures, email the authors.) In these grayscale images the darkest regions indicate modes of energy, possible targets. The 2<sup>nd</sup> and 3<sup>rd</sup> order MF processors show strong peaks at both the ship and towed beacon positions; there are more sidelobes around the CW source position for 3<sup>rd</sup> order. The 4<sup>th</sup> order MF processor shows a weak value at the ship position. This fact suggests that the 4<sup>th</sup> order component of ship noise is not as strong as its 2<sup>nd</sup> and 3<sup>rd</sup> order in the frequency band 73-133Hz in the MEVA3 trial.

For the higher frequency band of 150-270Hz, the 2<sup>nd</sup>, 3<sup>rd</sup> and 4<sup>th</sup> order MF processors are shown in Figs 6, 7 and 8, respectively. All figures show a strong peak at the ship and position, **but only the higher order shows significant relative energy from the deeper towed source.** Again, we have higher sidelobe level when MFP's order is increased. We note that the ship has strong components for not only second order but either white or non-white Gaussian random interferences since the third and higher odd moments and third and higher order cumulants of Gaussian random interferences are zero.

We have examined the higher order content of experimental ship and towed beacon data as it effects the MF processor for estimation of location using different frequency bands. For the frequency band 73-133Hz the 2<sup>nd</sup> and 3<sup>rd</sup> order MF processors show strong peaks at both the ship and towed beacon positions, but there are also more sidelobes around the CW source position for 3<sup>rd</sup> order than for lower order.

The 4<sup>th</sup> order MF processor at 73-133Hz shows a weak value at at both the ship and towed beacon positions. This fact suggests that the 4<sup>th</sup> order component of ship noise is not as strong as its 2<sup>nd</sup> and 3<sup>rd</sup> order in the frequency band 73-133Hz.

For the higher frequency band of 150-270Hz, the 2<sup>nd</sup>, 3<sup>rd</sup> and 4<sup>th</sup> order MF processors all show strong peaks at the ship position, however we again have higher sidelobes than for the lower order. This fact indicates that the ship has a strong component for not only second order but also third and fourth order in the frequency band 150-270 Hz in comparison with other sources in in the environment. **We note particularly that in this band, the lower source is more clearly identified by the higher order statistics, as it is relatively undetected by the 2<sup>nd</sup> order MFP compared to the primary target.**

#### 4. References

- [1] B. Kaufhold, R. L. Kirlin, R. M. Dizaji, "Blind System Identification Using Normalized Fourier Coefficient Gradient Vectors Obtained from Time-Frequency Entropy based Blind Clustering of Data Wavelets", *Signal processing, A Review Journal*, Vol. 9, pages 18-35, 1999
- [2] B. Baggeroer, W. A. Kuperman, P. N. Mikhalevsky, "An overview of matched field methods in ocean acoustics", *IEEE Journal of Oceanic Engineering*, Vol. 18, No. 4, pages 401-424, 1993.
- [3] E. K. Westwood, C. T. Tindle, N. R. Chapman, "A normal mode model for multi-layered acoustoelastic ocean environments based on an analytic reflection coefficient method", *J. Acoust. Soc. Amer.*, Vol. 95, page 2908, 1994.
- [4] H. Schmidt, "SAFARI: Seismo-acoustic fast field algorithm for range independent environments", User's guide, SR 113, *SACALANT ASW* Research center, La Spezia, Italy, 1987.
- [5] Mendel J. M., "Tutorial on higher-order statistics (spectra) in signal processing and system theory: theoretical results and some applications", *Proceedings of the IEEE*, Vol. 79, pp. 278-305, 1991
- [6] Nikias C. L., and Raghuvver M., "Bispectrum estimation: a digital signal processing framework", *Proceeding of IEEE*, Vol. 75, pp. 869-891, 1987
- [7] Pan R., Nikias C. L., "The complex cepstrum of higher-order moments", *IEEE Trans. On ASSP*, Vol. 36, pp. 186-205, 1988
- [8] Nikias, C.L., and A.P. Petropulu, *Higher-Order Spectra Analysis*, Prentice Hall, 1993.

# MULTIWINDOW BISPECTRAL ESTIMATION

Yngve Birkelund and Alfred Hanssen

University of Tromsø, Physics Department  
Electrical Engineering Group N-9037 Tromsø, Norway.  
E-mail: yngve@phys.uit.no, alfred@phys.uit.no

## ABSTRACT

The bispectral density provides crucial information about non-Gaussian and/or non-linear properties of stochastic processes. In practice however, bispectral estimators are prone to be inaccurate and statistically inconsistent. In this paper we have discussed the statistical properties of non-parametric direct bispectral estimators. Several multitaper based bispectral estimators are presented, including a recently developed approach giving better frequency resolution. We will show that the bias and variance of these estimators are governed mainly by a quantity we call the total bispectral window. Our conclusion is that classical bispectral estimation using biperiodogram in combination of tapering and/or bifrequency smoothing are outperformed by multitaper based bispectral estimators presented in this paper.

## 1. INTRODUCTION

It is well known that estimates of bispectra and other polyspectra are prone to be noisy and statistically inconsistent. The problems are particularly severe when small data sets are available, or when the data comes from a nonstationary process. In particular, the naive bispectral estimator (the so-called *biperiodogram*) is anti-consistent.

In 1989, Thomson suggested [1] a multitaper estimator for bispectral densities, by extending his established power-spectral estimation technique. Thomson gave a useful approximation of the variance of his multitaper bispectral estimator assuming Gaussian processes. Numerical verification of Thomson's approximative variance expression and numerical examples for non-Gaussian processes were provided by the present authors recently [2].

In this paper, we will discuss the statistical properties of non-parametric direct bispectral estimators, with special emphasis on multitaper estimators. In particular, we show that the bias of any non-parametric bispectral estimator is governed by a quantity we call the *total bispectral window*, and that the variance may be approximated by a term depending on the same quantity. We will also generalize the multitaper approach and present a recently developed bispectral estimator with better bias properties for rapidly varying bispectral densities [3]. We will briefly discuss leakage effects in bispectral estimation, and introduce the use of data adaptive weight functions to control these effects in the multitaper approach [4]. Finally, we discuss the applicability of multitaper based bispectral estimators.

## 2. SPECTRAL REPRESENTATION

In this paper we will assume that  $N$  samples are available, equally spaced in time with  $\Delta t = 1$ , from a real valued, zero-mean, stationary and ergodic stochastic process. Then there exist a Cramér spectral representation [5, 6]

$$x[n] = \int_{-1/2}^{1/2} \exp(j2\pi fn) d\mathbf{X}(f) \quad (1)$$

where  $x[n]$  for  $n = 0, 1, \dots, N-1$  are the data samples and  $d\mathbf{X}(f)$  is the increment process at frequency  $f$ . The relationship between available data represented by the standard Fourier transformed data  $X(f) = \sum_{n=0}^{N-1} x[n] \exp(-j2\pi fn)$  and the increment process  $d\mathbf{X}(f)$  can be written as [7]

$$X(f) = \int_{-1/2}^{1/2} \tilde{D}(f-f') d\mathbf{X}(f'). \quad (2)$$

Here  $\tilde{D}(f) = D(f) \exp[j(N-1)\pi f]$  is a phase-shifted version of the Dirichlet kernel  $D(f) = \sin(N\pi f)/\sin(\pi f)$ . From the properties of the Dirichlet kernel, it is easy to understand two fundamental properties of Fourier based estimators: First,  $D(f)$  is zero for the harmonic frequencies  $f = i/N$  for  $i = 1, 2, \dots, N-1$ . This implies that for white noise, we can obtain uncorrelated estimates of  $d\mathbf{X}(f)$  at any two different harmonic frequencies. Second, the Dirichlet kernel has a large sidelobe level which gives raise to severe spectral leakage effects [8, 9].

## 3. CONVENTIONAL NON-PARAMETRIC BISPECTRAL ESTIMATORS

Using the spectral representation for the stochastic process, the integrated bispectrum  $B(f_1, f_2) df_1 df_2$  is defined by [1, 6]

$$B(f_1, f_2) df_1 df_2 = \text{Cum}[d\mathbf{X}(f_1) d\mathbf{X}(f_2) d\mathbf{X}(f_3)] \quad (3)$$

where  $B(f_1, f_2)$  is the bispectral density, and  $f_1 + f_2 + f_3 = 0$ . Using the Fourier transform  $X(f)$  to approximate the increment process  $d\mathbf{X}(f)$ , the resulting (naive) bispectral estimator is the *biperiodogram*  $\hat{B}^{per}(f_1, f_2)$  given by

$$\hat{B}^{per}(f_1, f_2) = \frac{1}{N} X(f_1) X(f_2) X^*(f_1 + f_2) \quad (4)$$

where the asterisk denotes complex conjugation.

The statistical properties of the biperiodogram are described by [10, 6]. Using the relationship between the Fourier transform and the increment process, it is easy to show that the expected

value can be written as a two-dimensional convolution between the true bispectrum of the process  $B(f_1, f_2)$  and the *rectangular bispectral kernel*  $D(f_1, f_2)$

$$E[\hat{B}^{per}(f_1, f_2)] = \int_{-1/2}^{1/2} D(f_1 - f'_1, f_2 - f'_2) B(f'_1, f'_2) df'_1 df'_2, \quad (5)$$

where  $D(f_1, f_2)$  can be expressed by means of the Dirichlet kernel as  $D(f_1, f_2) = \tilde{D}(f_1)\tilde{D}(f_2)\tilde{D}^*(f_1 + f_2)/N = D(f_1)D(f_2)D(f_1 + f_2)/N$ . Asymptotically ( $N \rightarrow \infty$ ) the kernel approaches a two-dimensional Dirac delta function making the biperiodogram asymptotically unbiased. For finite  $N$ , however, the rectangular bispectral kernel implies a leakage in the bifrequency domain.

Assuming a Gaussian process, and that  $X(f_1)$ ,  $X(f_2)$  and  $X(f_1 + f_2)$  are uncorrelated, the variance has been approximated by [10, 6]

$$\text{Var}[\hat{B}^{per}(f_1, f_2)] \simeq NS(f_1)S(f_2)S(f_1 + f_2), \quad (6)$$

where  $S(f)$  is the true power spectrum of the process, and  $f_1 \neq (0, \pm 1/2)$ ,  $f_2 \neq (0, \pm 1/2)$  and  $|f_1 + f_2| \neq (0, \pm 1/2)$ . From eq. (6) it is clear that the biperiodogram is anti-consistent, since the variance increases as the number of data samples  $N$  increases. This anti-consistency is certainly not acceptable for an estimator, and the raw biperiodogram should therefore be avoided in general.

### 3.1. Frequency smoothing

The variance of the biperiodogram can obviously be reduced by frequency smoothing in the bispectral domain. The covariance of the biperiodogram shows that different frequency pairs  $(f_k, f_l) \neq (f_m, f_n)$  are uncorrelated for harmonic frequencies  $f_i = i/N$ ;  $i = 0, \pm 1, \dots, \pm N-1$  [10, 6]. Applying a discrete *bispectral smoothing window*  $G(f_1, f_2)$  of the biperiodogram, can therefore reduce the variance at the cost of poorer frequency resolution.

The so-called uniform smoothing window [11] has a constant value within a hexagonal bifrequency region of support. The size of the hexagon is user specified, and is controlled by a single integer parameter  $a$  describing the frequency smoothing bandwidth (see [2] for more details). To simplify the discussion of conventional non-parametric estimators, we will restrict ourselves to the uniform smoothing window in the rest of this paper.

Using the assumption that the bispectral density of the process is approximately constant within the smoothing bandwidth, the uniform smoothing does not change the expectation value of the biperiodogram. Since the biperiodogram at pairs of harmonic frequencies are uncorrelated, it is easy to show that uniform smoothing reduces the variance approximately by the factor  $1/C$ , the number of non-zero points in  $G(f_1, f_2)$ , for a white Gaussian process.

### 3.2. Tapering

Tapering is the well known solution for reducing spectral leakage in power spectral estimation. Denoting the data taper by  $v[n]$  and the available data by  $x[n]$ , the tapered data  $y[n]$  is obtained by  $y[n] = x[n]v[n]$ , for  $n = 0, 1, \dots, N-1$ . The effect on the expectation value, can easily be seen using the relationship between Fourier transformed tapered data  $Y(f)$  and the true increment process  $dX(f)$ ,

$$Y(f) = \int_{-1/2}^{1/2} V(f - f') dX(f') \quad (7)$$

Here the convolution kernel  $V(f)$  is the discrete Fourier transform of the data taper  $v[n]$ . With the use of standard data tapers as the Hanning taper, the kernel in eq. (7) will be modified to have a broader mainlobe and lower sidelobe level than the Dirichlet kernel in eq. (2) [8]. Leakage is thus reduced at the expense of a poorer frequency resolution. If the taper is normalized by  $\sum_{n=0}^{N-1} v^2[n] = N$ , a tapered biperiodogram can be obtained using  $Y(f)$  instead of  $X(f)$  in eq. (4).

The statistical properties of tapered biperiodograms are closely connected to those of the biperiodogram discussed above. The expected value can be written as a convolution between the *tapered bispectral kernel*  $V(f_1, f_2)$  and the true bispectrum, as in eq. (5), where  $V(f_1, f_2) = V(f_1)V(f_2)V^*(f_1 + f_2)$ . This means that the bispectral leakage can be reduced because of lower sidelobe level in  $V(f_1, f_2)$ , but the use of tapering also reduces the frequency resolution since the mainlobe of  $V(f_1, f_2)$  is wider than the rectangular bispectral kernel.

### 3.3. Tapering and frequency smoothing

The use of tapering in combination with frequency smoothing introduces some properties that are difficult to quantify in the resulting bispectral estimate. Using the approach in [9] (pp. 243-246), it is possible to show that the expected value of a frequency smoothed tapered biperiodogram can be written as a convolution between the true bispectrum and the *total bispectral window*  $W(f_1, f_2)$  given by

$$W(f_1, f_2) = \int_{-1/2}^{1/2} G(f_1 - f'_1, f_2 - f'_2) V(f'_1, f'_2) df'_1 df'_2 \quad (8)$$

where  $G(f_1, f_2)$  is the bispectral smoothing window and  $V(f_1, f_2)$  is the tapered bispectral kernel.

The variance properties for this tapered and smoothed estimator, are somewhat complicated. Use of tapers other than the rectangular, will introduce correlation between the bifrequency  $(f_1, f_2)$  and its surroundings even if  $f_i$  are harmonic frequencies. This implies less effective frequency smoothing regarding variance reduction [12]. The total frequency smoothing area given by the  $W(f_1, f_2)$  is, however, slightly broader with use of other tapers than the rectangular. The bifrequency smoothing effect together with the less effective use of data, makes the variance of tapered and smoothed biperiodogram larger when tapers other than the rectangular is used.

## 4. OPTIMAL DATA TAPERS

While conventional non-parametric bispectral estimation seems as an ad-hoc combination of tapering and smoothing, the multitaper approach is a result of a strict optimization criterion. Maximizing the spectral concentration

$$\lambda = \frac{\int_{-f_B}^{f_B} |V(f)|^2 df}{\int_{-1/2}^{1/2} |V(f)|^2 df} \quad (9)$$

where  $V(f)$  is the Fourier transform of the taper  $v[n]$  and  $f_B$  is a chosen bandwidth, leads to a set of orthonormal data tapers known as the Slepian tapers, or Discrete Prolate Spheroidal Sequences (DPSS). It is easy to show that the maximization of  $\lambda$  in eq. (9) leads to an  $N$ th-order eigenvalue problem in the time domain [7]

$$\mathbf{A}\mathbf{v} = \lambda\mathbf{v}, \quad (10)$$

where  $\mathbf{v} = [v[0]v[1] \dots v[N-1]]^T$  and the matrix  $\mathbf{A}$  has elements  $A_{n_1, n_2} = \sin[2\pi f_B(n_1 - n_2)]/\pi(n_1 - n_2)$ . The eigenvectors of this eigenvalue problem are the DPSS tapers, which we denote as  $v_k[n]$  and order by decreasing corresponding spectral concentration  $\lambda_k$ . Note that the tapers are orthonormal and that their Fourier transforms are doubly orthogonal [7].

Slepian showed that the spectral concentration  $\lambda_k$  is close to unity for tapers  $v_k[n]$  for orders  $k = 0, 1, \dots, 2Nf_B - 1$ , where  $K = 2Nf_B$  is known as the Shannon number, and that it falls rapidly towards zero for orders beyond  $K-1$ . The optimal number of DPSS to be used in a multitaper approach is therefore  $K = 2Nf_B$ , and the resulting *total spectral window*

$$W(f) = \sum_{k=0}^{K-1} \lambda_k |V_k(f)|^2 / \sum_{k=0}^{K-1} \lambda_k \quad (11)$$

will approximate an ideal band limited filter [9]. The highest order DPSS taper used has the lowest spectral concentration, and is therefore the taper with the highest sidelobe level. Reducing the number  $K$  of tapers in the multitaper approach results in lower sidelobes in the total spectral window.

Assuming a predefined peaked spectrum prototype shape  $S_x(f)$ , it is possible to obtain a set of orthonormal peak matched (PM) tapers that have better frequency resolution than the ideal flat smoothing in DPSS tapers from eq. (10). With slightly different notation than in [13], we will assume a logarithmic triangular spectral peak with 0 dB in  $f = 0$ ,  $C$  dB in  $f = \pm f_B$  and  $-\infty$  dB outside the half-bandwidth  $f_B$  as in the case of DPSS. The PM tapers are the solutions of the eigenvalue problem

$$\mathbf{P}\mathbf{v} = \lambda\mathbf{v}, \quad (12)$$

with  $v_k[n]$  as eigenvectors and corresponding eigenvalues  $\lambda_k$ . As for the DPSS case, we will order the eigenvalues in decreasing order and use the  $K = 2Nf_B$  lowest order PM tapers in our multitaper approach. The Toeplitz covariance matrix  $\mathbf{P}$  has the elements  $P_{n_1, n_2} = r_x[n_1 - n_2] * \sin(2\pi f_B(n_1 - n_2))/\pi[n_1 - n_2]$ , where  $r_x[n]$  is the covariance sequence corresponding to  $S_x(f)$  and  $*$  denotes a convolution. The resulting spectral window using PM tapers will approximate the predefined spectrum prototype.

The Fourier transform  $V_k(f)$  of the PM tapers has approximately the same sidelobe level for any order  $k$ , so we cannot reduce the sidelobe in the total spectral window by using fewer tapers as in the DPSS taper case. To decrease the effect of leakage, we therefore have to introduce a frequency selective penalty spectrum  $S_g(f)$  in the eigenvalue problem [13]

$$\mathbf{P}\mathbf{v} = \lambda\mathbf{M}\mathbf{v}. \quad (13)$$

Here, the matrix  $\mathbf{M}$  has a Toeplitz structure with elements  $M_{n_1, n_2} = r_g[n_1 - n_2]$ , where  $r_g[n]$  is the covariance sequence corresponding to the penalty spectrum  $S_g(f)$ . The penalty spectrum has a flat response of 0 dB inside the chosen bandwidth  $f_B$ , and a level of  $G$  dB outside. The resulting PM tapers from the generalized eigenvalue problem in eq. (13), have  $G$  dB lower sidelobe level at the cost of even faster decreasing eigenvalues and thereby less effective number of tapers used in the total spectral window. Note that by choosing  $G = 0$ , the generalized eigenvalue problem in eq. (13) is reduced to eq. (12), so ordinary PM tapers are actually PM tapers without sidelobe suppression. Note also that the DPSS tapers can be obtained from eq. (13) by choosing  $C = G = 0$  [13].

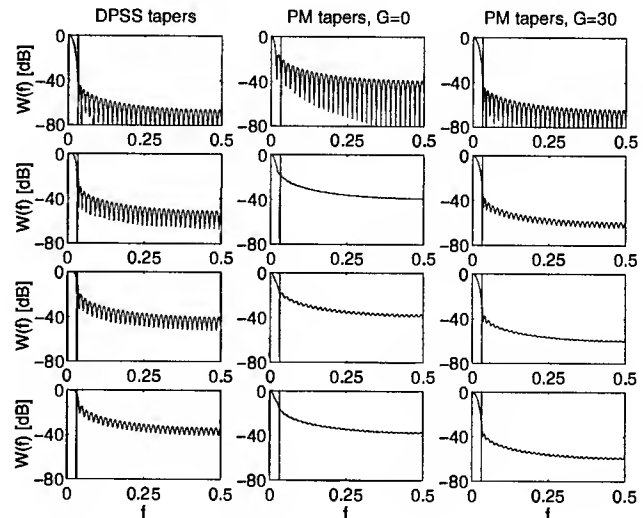


Figure 1: Total spectral window for three different sets of orthogonal tapers in the case of  $N = 64$ ,  $f_B = 2/N$ . Left: DPSS tapers; Middle: PM tapers with  $C = -20$  and  $G = 0$ ; Right: PM tapers with  $C = -20$  and  $G = 30$ .

To summarize, the use of tapers that are solutions of the generalized eigenvalue problem in eq. (13) gives a controlled frequency smoothing effect in the spectral domain. In Fig. 1 we show the total spectral window for three different sets of tapers for the case  $N = 64$  and  $f_B = 2/N$ . The left panel shows the DPSS tapers, the middle panel shows the PM tapers with  $C = -20$  dB and  $G = 0$ , and the right panel shows the PM tapers with  $C = -20$  dB and  $G = 30$  dB. The total spectral window is plotted as a function of number of tapers used, increasing from  $K = 1$  to  $K = 4$  from top to bottom. The corresponding eigenvalues for the DPSS taper and PM tapers with and without sidelobe suppression are shown in Table 1. The variance reduction can be connected to the effective number of orthonormal tapers actually in use, and is therefore closely connected to the corresponding set of eigenvalues. The DPSS tapers have the best variance properties since all eigenvalues are close to unity, while the peak matched tapers offer a frequency selective multitaper approach at the expense of variance reduction.

Order $k$	0	1	2	3
DPSS	0.9999	0.9976	0.9596	0.7220
PM, $G = 0$	0.5363	0.2057	0.0792	0.0297
PM, $G = 30$	0.4218	0.0483	0.0020	0.0001

Table 1: Eigenvalues  $\lambda_k$  for DPSS tapers and PM tapers with and without sidelobe suppression. The four lowest order eigenvalues are shown for the case  $N = 64$ ,  $f_B = 2/N$  and  $C = -20$ .

#### 4.1. Adaptive weight functions

To reduce spectral leakage in the multitaper power spectral estimator, Thomson introduced weight functions  $d_k(f)$  for each order of taper to obtain improved estimates of  $d\mathbf{X}_k(f)$  [7]. Using a robust adaptive approach, DPSS tapers with high sidelobe level are

down-weighted in frequency regions where leakage can influence the estimate. Since each taper is considered individually for leakage, the adaptive approach also eliminates the need for choosing the optimal number of tapers  $K \leq 2Nf_B$  in the multitaper estimate. More details concerning about the adaptive approach for determining these weight function can be found in [7, 4].

The use of data adaptive weight functions  $d_k(f)$  in bispectral estimation have been thoroughly discussed in [4]. The effect of leakage in bispectral estimation are more complicated than for the power spectral estimation case. In brief, leakage can strongly influence the estimator variance while the bias can be negligible.

To detect frequencies where leakage can influence the estimation of  $d\mathbf{X}(f)$ , the adaptive approach requires tapers where leakage is not present. The lowest order of DPSS tapers has the lowest possible sidelobe level for the chosen bandwidth, providing *leakage free* estimates of the true increment process. In the DPSS taper case, these weight functions therefore effectively reduces the leakage. For the peak matched tapers the adaptive approach is useless since tapers for all order has approximately the same sidelobe level.

## 5. MULTITAPER BISPECTRAL ESTIMATORS

Let  $x[n]$  be the data available for  $n = 0, 1, \dots, N-1$ . Using any orthogonal set of basis functions  $v_k[n]$  and corresponding eigenvalue  $\lambda_k$  for  $k = 0, 1, \dots, K-1$ , we obtain a set of tapered data as  $y_k[n] = x[n]v_k[n]$  with corresponding Fourier transform  $Y_k(f)$ .

The multitaper approach has been applied to bispectral estimators in [7, 1, 2, 4]. A general approach for multitaper bispectral estimation (MBE) can be written as a weighted sum of all combinations of individually tapered biperiodograms

$$\hat{B}(f_1, f_2) = \frac{1}{U_3} \sum_{k,l,m=0}^{K-1} Q(k, l, m) \hat{B}_{(k,l,m)}(f_1, f_2) \quad (14)$$

where the tapered biperiodogram of order  $(k, l, m)$  is

$$\hat{B}_{(k,l,m)}(f_1, f_2) = Y_k(f_1)Y_l(f_2)Y_m^*(f_1 + f_2) \quad (15)$$

and the three-dimensional weighting function given by

$$Q(k, l, m) = \sqrt{\lambda_k \lambda_l \lambda_m} \sum_{n=0}^{N-1} v_k[n]v_l[n]v_m[n]. \quad (16)$$

The normalization constant  $U_3$  is defined by

$$U_3 = \sum_{k,l,m=0}^{K-1} Q^2(k, l, m) / \sqrt{\lambda_k \lambda_l \lambda_m} \quad (17)$$

to ensure that the bispectral estimator is unbiased for white noise.

The use of data adaptive weight functions  $d_k(f)$  in bispectral estimation modifies the three-dimensional weight function in eq. (16) to be bifrequency selective

$$Q'_{(k,l,m)}(f_1, f_2) = Q_{(k,l,m)} d_k(f_1) d_l(f_2) d_m(f_1 + f_2). \quad (18)$$

To obtain an unbiased bispectral estimate, the new normalization constant  $U'_3$  also depends on bifrequencies  $(f_1, f_2)$

$$U'_3(f_1, f_2) = \sum_{k,l,m=0}^{K-1} Q^2_{(k,l,m)} d_k(f_1) d_l(f_2) d_m(f_1 + f_2). \quad (19)$$

The modification caused by  $d_k(f)$  in eq. (18) and eq. (19) is sufficient to make the MBE in eq. (14) resistant against leakage.

### 5.1. Statistical properties

We have examined the statistical properties of the MBE based estimators in great detail. In the following, we will discuss our findings in some detail.

The expectation value of the general multitaper bispectral in eq. (14) can be shown to be

$$E[\hat{B}(f_1, f_2)] = \int_{-1/2}^{1/2} W(f_1 - f'_1, f_2 - f'_2) B(f'_1, f'_2) df'_1 df'_2 \quad (20)$$

where the *total bispectral window* [2] is given by

$$W(f_1, f_2) = \frac{1}{U_3} \sum_{k,l,m=0}^{K-1} Q(k, l, m) W_{k,l,m}(f_1, f_2) \quad (21)$$

and the *bispectral window of order*  $(k, l, m)$  is given by

$$W_{k,l,m}(f_1, f_2) = V'_k(f_1) V'_l(f_2) [V'_m(f_1 + f_2)]^* \quad (22)$$

Assuming Gaussian data and distinct frequencies  $f_1, f_2$  and  $f_1 + f_2$ , the smoothing effect of the true bispectral density in the MBE leads to a variance decrease. Only considering this smoothing effect, the variance of the MBE can be approximated by

$$\text{var}[\hat{B}(f_1, f_2)] \simeq \int_{-1/2}^{1/2} W^2(f'_1 - f_1, f'_2 - f_1) \text{var}\{\hat{B}^{per}(f'_1, f'_2)\} df'_1 df'_2. \quad (23)$$

Here  $\text{var}\{\hat{B}^{per}(f'_1, f'_2)\} = NS(f_1)S(f_2)S(f_1 + f_2)$  is the asymptotic variance of the biperiodogram [10]. This implies that the bispectral estimate is consistent for fixed  $f_B$ , since asymptotically  $\text{var}[\hat{B}(f_1, f_2)] = 0$ . For conventional bispectral estimation, the tapering in combination with bifrequency smoothing implies a *variance increase* compared to only smoothing of the biperiodogram since tapering implies less effective use of data. For multitapering the effective use of data is better than for a single taper [9], so the difference between the approximation in eq. (23) and the true variance is small.

Statistical properties of the adaptive MBE are hard to obtain in general since the calculation of weight functions depends on the process in study. Processes with small dynamical range in the true increment process  $d\mathbf{X}(f)$  have no long range leakage, and the adaptive MBE are therefore close to the MBE with all  $K = 2Nf_B$  tapers used. For processes with large dynamical range in  $d\mathbf{X}(f)$ , the total bispectral window  $V(f_1, f_2)$  must be redefined to also depend on the actual bifrequency  $(f_1, f_2)$  under study,

$$V(f_1, f_2, f'_1, f'_2) = \sum_{k,l,m=0}^{K-1} Q'_{(k,l,m)}(f'_1, f'_2) V_{(k,l,m)}(f_1, f_2) \quad (24)$$

where  $V_{(k,l,m)}(f_1, f_2) = V_k(f_1) V_l(f_2) V_m(f_1 + f_2)$ . For bispectral regions where the magnitude is low, the weight function will down-weight the tapers with high sidelobe levels so leakage is avoided. This down-weighting of tapers will reduce the variance of the adaptive MBE in lower bispectral parts, since leakage from these tapers contributes to the variance.

Extensive Monte Carlo simulations of the non-parametric estimators discussed in this paper have verified our results on the statistical properties [2, 3, 4].



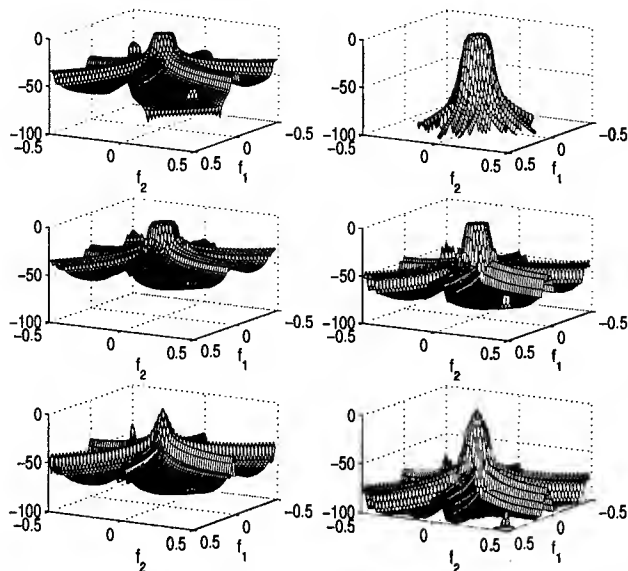


Figure 2: Total bispectral window for three classes of bispectral estimators. Upper left: Uniformly smoothed biperiodogram. Upper right: Hanning tapered and uniformly smoothed biperiodogram. Middle left: MBE using  $K = 2Nf_B$  DPSS tapers. Middle right: MBE using  $K = 2Nf_B - 3$  DPSS tapers. Lower left: MBE using peak matched tapers. Lower right: MBE using peak matched tapers with suppressed sidelobes. All bispectral windows are for the case:  $N = 64$ ,  $Nf_B = a = 4$ ,  $C = -20$  dB and  $G = 30$  dB.

## 5.2. Total bispectral windows

The total bispectral windows are plotted in Fig. 2 for some of the non-parametric bispectral estimators discussed in this paper. These examples have approximately the same hexagonal region of support.

Use of Hanning taper in combination of uniform smoothing (upper right) will lower the sidelobe level significantly compared to the use of a rectangular taper (upper left), but the support in the total bispectral window also is rounded and slightly wider.

The MBE using  $K = 2Nf_B$  DPSS tapers (middle left) have approximately the same sidelobe level as the biperiodogram, but the "edges" going out from support are lower. Using fewer DPSS tapers in the MBE (middle right) will lower the sidelobe level without destroying the flat support.

The "pyramidal" support in the peak matched taper case with (lower left) and without (lower right) sidelobe suppression clearly differs from the other total bispectral windows. The linear decay (in dB) to  $-20$  dB at the edge of hexagonal support, means that these tapers provide better frequency resolution. This is achieved, as usual in spectral estimation, at the cost of higher variance.

## 6. CONCLUSION

While the expected value and bias are completely described by the total bispectral window alone, we also have to consider the effective use of data to describe the variance properties.

The statistical performance for a specific estimator will of course depend on the particular process under study. For slowly varying bispectra, Thomson's original multitaper approach using

DPSS tapers is the best choice. If the process has large dynamical range in the true increment process, we have to reduce the leakage to get satisfactory results. For non-parametric estimation this means data tapering, where the use of frequency selective weight functions in combination of DPSS tapers seem to be an obvious choice. Conventional techniques with tapered and smoothed biperiodogram implies less effective use of data, and thus have higher variance.

The use of peak matched tapers provides a good combination of low bias and variance reduction in the MBE for rapidly varying bispectra. In cases of large dynamical range in the true increment process, we conclude that peak matched tapers with suppressed sidelobes should be applied.

## REFERENCES

- [1] D. J. Thomson, "Multi-window bispectrum estimates," *Proc. IEEE Workshop on Higher-Order Statistics*, Vail, Colorado, pp. 19–23, 1989.
- [2] Y. Birkelund and A. Hanssen, "Multitaper estimators for bispectra," *Proc. IEEE Workshop on Higher-Order Statistics*, Caesarea, Israel, pp. 207–211, 1999.
- [3] Y. Birkelund and A. Hanssen, "Bispectral estimation: A multitaper approach," *Proc. X European Signal Processing Conference*, Tampere, Finland, (in press), 2000.
- [4] Y. Birkelund and A. Hanssen, "Adaptive bispectral estimation using Thomson's multitaper approach," *Proc. IEEE Symposium on Adaptive Systems for Signal Processing, Communication and Control*, Lake Louise, Alberta, Canada, (in press), 2000.
- [5] M. B. Priestley, *Spectral analysis and time series*, Academic press, 1987.
- [6] M. Rosenblatt, *Stationary sequences and random fields*, Birkhäuser, 1985.
- [7] D. J. Thomson, "Spectrum estimation and harmonic analysis," *Proc. IEEE*, vol. 70, no. 9, pp. 1055–1096, 1982.
- [8] F. J. Harris, "On the use of windows for harmonic analysis with the discrete fourier transform," *Proc. IEEE*, vol. 66, no. 1, pp. 51–83, 78.
- [9] D. B. Percival and A. T. Walden, *Spectral analysis for physical applications: Multitaper and conventional univariate techniques*, Cambridge, 1993.
- [10] D. R. Brillinger and M. Rosenblatt, "Asymptotic theory of estimates of k-th order spectra," *Spectral analysis of time series*, pp. 153–188, 1967.
- [11] C. L. Nikias and M. R. Raghuveer, "Bispectrum estimation: A digital signal processing framework," *Proc. IEEE*, vol. 75, no. 7, pp. 869–891, 1987.
- [12] P. J. Huber, B. Kleiner, T. Gasser, and G. Dumermuth, "Statistical methods for investigating phase relations in stationary stochastic processes," *IEEE Trans. Audio Electroacoust.*, vol. AU-19, no. 1, pp. 78–86, 1971.
- [13] M. Hansson and G. Salomonsen, "A multiple window methods for estimation of peaked spectra," *IEEE Trans. Signal Processing*, vol. 45, no. 3, pp. 778–781, 1997.



# GLOBAL CONVERGENCE OF A SINGLE-AXIS CONSTANT MODULUS ALGORITHM

A. Shah, S. Biracree, R. A. Casas, T. J. Endres, S. Hulyalkar, T. A. Schaffer, C. H. Strolle

NxtWave Communications

One Summit Square

Langhorne, PA 19047

{anand, biracree, raul, endres, samirh, schaffer, cstrolle}@nxtwavecomm.com

## ABSTRACT

We propose a modification to the Constant Modulus Criterion for real valued sources processed with a complex valued receiver. Our modification is called Single-Axis Constant Modulus Criterion (SA-CM) because it operates solely on the real component of the complex equalizer output. We show that under idealized conditions, a finite length, baud-spaced, complex valued equalizer minimizing the SA-CM criterion admits only desirable global minima settings that are ISI-free. A single-axis receiver architecture is compared to other receiver architectures for real valued sources and staggered modulation schemes. Simulation examples using vestigial sideband (VSB) signaling verify our methods.

## 1. INTRODUCTION

Modern digital receivers often rely on blind equalization techniques to mitigate unknown channel distortions. Blind methods are desirable because they do not rely on a periodic transmission of a training sequence, thus increasing data throughput and allowing for equalizer adaptation at every symbol instance. The Constant Modulus Algorithm (CMA) is a popular blind equalization technique used in high data-rate applications due to its robustness under practical signaling conditions [4].

It is often the case that a digital receiver uses complex valued signal processing even though the data source is real valued or encodes original information into only one dimension. Complex signal processing may be required by the receiver since synchronization and equalization functions operate on passband data that is not precisely downconverted to baseband. Most treatments of CMA assume either real valued signal processing for a real valued source (such as PAM) or complex valued signal processing for complex valued sources (such as QAM) - [7] is an exception. In [7], Papadakis shows that because a BPSK source is not circularly symmetric, *i.e.*  $E\{s^2\} \neq 0$ , the Constant Modulus (CM) cost function admits global minima settings that result in a closed-eye combined channel-equalizer response.

We present a modification to the CM criterion appropriate for real valued sources that are processed by complex valued receivers, in which equalizer coefficients are updated using real-part extraction of the equalizer filter result. We show that a finite length, baud-spaced, complex valued equalizer minimizing the SA-CM criterion ad-

mits only minima that are global and result in open-eye settings, thus excluding the undesirable settings described in [7]. As fractionally-spaced equalizers exploit temporal diversity, single-axis equalizers exploit phase diversity in complex valued channels.

Tu [10] applies a similar concept for staggered modulation formats (such as staggered-QAM and vestigial sideband modulation) using minimum mean square error (MMSE) equalization. In staggered modulation, information is encoded independently onto in-phase (I) and quadrature-phase (Q) carriers, with the carriers staggered in time by typically half the symbol period relative to standard QAM. Reference [10] shows that alternatively minimizing the Mean Square Error (MSE) over I and Q samples results in a lower MSE performance than minimizing the complex valued error term. Through example, we show that SA-CMA is additionally applicable to staggered modulation formats.

The next section describes a communication model using real valued data sources with complex valued signal processing and provides motivation for employing a complex valued receiver. Section 3 introduces the SA-CM criterion and shows its perfect symbol recovery properties. Section 4 provides simulation examples and applies SA-CMA for staggered modulations. Section 5 provides concluding remarks. Section 6 provides directions of future work.

## 2. RECEIVER ARCHITECTURES FOR REAL VALUED SOURCES

We begin with a communication model of a real valued, sub-Gaussian (*i.e.*  $E\{s^4(n)\} < 3E\{s^2(n)\}^2$ ), zero-mean, i.i.d. source  $\{s(n)\}$ ,  $n \in \mathbb{Z}$ ,  $s(n) \in \mathbb{R}$ . This real valued source is filtered through a complex valued FIR channel described in matrix notation as

$$\mathbf{C} = \begin{pmatrix} c_0 & & & & \\ c_1 & c_0 & & & \\ \vdots & c_1 & \ddots & & \\ c_{N_c} & \vdots & & c_0 & \\ & c_{N_c} & & c_1 & \\ & & \ddots & \vdots & \\ & & & & c_{N_c} \end{pmatrix}$$

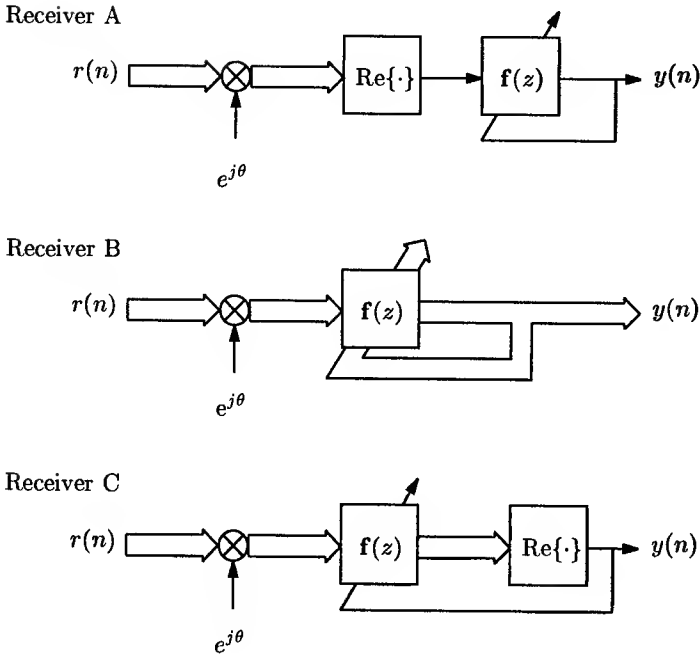


Figure 1: Receiver A: real valued equalizer  $\mathbf{f}(z)$  updated with real valued estimates  $y(n)$ . Receiver B: complex valued equalizer  $\mathbf{f}(z)$  updated with complex valued estimates  $y(n)$ . Receiver C: complex valued equalizer  $\mathbf{f}(z)$  updated with real valued estimates  $y(n)$ .

with  $c_i \in \mathbb{C}$ . We present for examination three receiver architectures, shown in Figure 1. Fat arrows denote complex valued signals; thin arrows denote real valued signals.

Each receiver consists of carrier phase correction of complex valued received samples followed by an adaptive FIR equalizer. Receiver A has a real valued equalizer operating on real valued data. Receiver B shows a complex valued equalizer that is updated according to complex valued estimates of the source symbols. As noted in [7], this receiver can result in a closed-eye combined channel-equalizer setting when the CM criterion is minimized.

Receiver C reflects the architecture we have termed as single-axis equalization, which employs a baseband, complex valued, baud-spaced FIR equalizer  $\mathbf{f} = (f_0, \dots, f_{N_f})^T$ ,  $f_i \in \mathbb{C}$  to generate real valued estimates  $y(n)$  of the source symbols. These estimates are given by

$$y(n) = \text{Re} \{ \mathbf{r}^T(n) \mathbf{f} \}$$

where the received signal is given by  $\mathbf{r}(n) = \mathbf{C}^T \mathbf{s}(n) + \mathbf{w}(n)$ , and  $\mathbf{s}(n) = (s(n), \dots, s(n - N_h))^T$ , with  $N_h = N_c + N_f$ , and  $\theta$  is assumed to be zero. (The effect of a non-zero phase offset,  $\theta$ , will be discussed in Section 4.2.) Note that  $\mathbf{w}(n) = (w(n), \dots, w(n - N_f))^T$ , where  $w(n)$  is an additive white Gaussian noise process.

Because  $\text{Im} \{s(n)\} = 0$ , we can rewrite the single-axis

equalizer output as

$$\begin{aligned} y(n) &= \text{Re} \{ \mathbf{r}^T(n) \} \text{Re} \{ \mathbf{f} \} - \text{Im} \{ \mathbf{r}^T(n) \} \text{Im} \{ \mathbf{f} \} \\ &= \mathbf{s}^T(n) (\text{Re} \{ \mathbf{C} \} \text{Re} \{ \mathbf{f} \} - \text{Im} \{ \mathbf{C} \} \text{Im} \{ \mathbf{f} \}) + \\ &\quad \text{Re} \{ \mathbf{w}^T(n) \mathbf{f} \} \end{aligned}$$

In this form, the single-axis equalizer is a linear, real valued, multi-channel receiver with sub-channels  $\text{Re} \{ \mathbf{C} \}$ ,  $-\text{Im} \{ \mathbf{C} \}$  and corresponding sub-filters  $\text{Re} \{ \mathbf{f} \}$ ,  $\text{Im} \{ \mathbf{f} \}$ . This channel-equalizer model is mathematically equivalent to the over-sampled channel-equalizer system in [2] or the antenna-array scheme proposed in [8]. Single-axis equalization exploits the channel phase diversity inherent in complex valued communication models using real valued sources. In the next section, we apply the properties of multi-rate and multi-channel systems to single-axis equalization and show the globally convergent behavior of SA-CMA.

### 3. SINGLE-AXIS CM CRITERION

We now show that a baud-spaced, finite length equalizer employing an adaptation strategy based on the SA-CM criterion can achieve perfect symbol recovery. To simplify notation, we use superscript notation  $(I)$  and  $(Q)$  to indicate real and imaginary components, respectively. Representing the channel with a matrix that isolates real and imaginary sub-channels, we have

$$\bar{\mathbf{C}} = \begin{pmatrix} c_0^{(I)} & & & -c_0^{(Q)} \\ c_1^{(I)} & & & -c_1^{(Q)} \\ & \ddots & & \vdots \\ c_{N_c}^{(I)} & & c_0^{(I)} & -c_{N_c}^{(Q)} \\ & & c_1^{(I)} & -c_1^{(Q)} \\ & & \vdots & \vdots \\ & & c_{N_c}^{(I)} & -c_{N_c}^{(Q)} \end{pmatrix}$$

In the absence of noise, the equalizer output can be written as  $y(n) = \mathbf{s}^T(n) \bar{\mathbf{C}} \bar{\mathbf{f}}$ , where

$$\bar{\mathbf{f}} = (f_0^{(I)}, f_1^{(I)}, \dots, f_{N_f}^{(I)}, f_0^{(Q)}, f_1^{(Q)}, \dots, f_{N_f}^{(Q)})^T.$$

The equalizer coefficients are obtained by minimizing the SA-CM criterion

$$J_{\text{sa-cm}}(\bar{\mathbf{f}}) = E \{ (y^2(n) - \gamma)^2 \}, \quad \gamma = \frac{E \{ s^4(n) \}}{E \{ s^2(n) \}}.$$

Equalizer coefficients are thus updated according to

$$\bar{\mathbf{f}}(n+1) = \bar{\mathbf{f}}(n) - \mu \bar{\mathbf{r}}(y^2(n) - \gamma) y(n)$$

where  $\bar{\mathbf{r}} = (r_0^{(I)}, r_1^{(I)}, \dots, r_{N_f}^{(I)}, -r_0^{(Q)}, -r_1^{(Q)}, \dots, -r_{N_f}^{(Q)})^T$ .

References [2] for a real valued source and [6] for a complex valued source, show that under a certain set of conditions, the CM criterion for multi-channel systems exhibits only global minima, and that these minima achieve perfect equalization (i.e. the combined channel-equalizer impulse response is a pure delay within a phase shift). The perfect equalizability conditions for a real valued source and channel-equalizer are rewritten here.

- (C1) No additive channel noise. (*i.e.*  $w(n) = 0, \forall n$ )
- (C2) Full row-rank channel matrix  $\bar{\mathbf{C}}$ . This necessitates the absence of zeros common to all sub-channel polynomials and sufficient filter length.
- (C3) Zero-mean, independent, and identically distributed source.
- (C4) Sub-Gaussian source (*i.e.*  $E\{s^4(n)\} < 3E\{s^2(n)\}$  in the real valued case).

Conditions (C1), (C3), (C4) are satisfied by our communication model. Condition (C2) requires that I and Q sub-channels are coprime, and that  $N_f \geq N_c - 1$ . Assuming these conditions are satisfied, the SA-CM cost surface exhibits the same global minima as an equivalent real valued dual-channel CM criterion.

Our communication system model with SA-CM criterion satisfies conditions (C1) to (C4), so that the single-axis CM criterion admits only desirable global minima. Hence, the equalizability condition for complex valued communication models, *i.e.*  $E\{s^2(n)\} \neq 0$ , does not apply to SA-CMA, and the undesirable equalizer settings proposed by Papadias [7] are not admitted due to real-part extraction. Note that SA-CM criterion is globally convergent with a finite length, baud-spaced equalizer by exploiting the phase diversity in the complex valued channel. This result is analogous to the global convergence result of the CM criterion in [2] and [6] using a finite length fractionally-spaced equalizer which exploits temporal diversity.

## 4. EXAMPLES

### 4.1. Two-tap channel

Consider transmission of a real valued source over a two-tap, complex valued channel  $c(z) = c_0 + c_1 z^{-1}$ . The output of our single-axis equalizer, using a single complex valued scalar filter  $f_0$  is given by

$$\begin{aligned} y(n) &= \text{Re} \{ f_0 (c_0 + c_1 z^{-1}) s(n) \} \\ &= (c_0^{(I)} f_0^{(I)} - c_0^{(Q)} f_0^{(Q)}) s(n) + \\ &\quad (c_1^{(I)} f_0^{(I)} - c_1^{(Q)} f_0^{(Q)}) s(n-1) \end{aligned}$$

It is possible to design  $f_0$  to recover a delayed version of the source  $y(n) = s(n - \delta)$ ,  $\delta = 0, 1$ , by solving the following set of equations

$$\begin{pmatrix} c_0^{(I)} & -c_0^{(Q)} \\ c_1^{(I)} & -c_1^{(Q)} \end{pmatrix} \begin{pmatrix} f_0^{(I)} \\ f_0^{(Q)} \end{pmatrix} = \begin{pmatrix} 1 \\ 0 \end{pmatrix} \quad \text{or} \quad \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

Figure 2 shows the SA-CM cost function for a BSPK source, *i.e.*  $s(n) \in \{\pm 1\}$  and  $c(z) = (0.3 + 0.4j) + (-0.1 + 0.2j)z^{-1}$  in (I,Q)-parameter space. Notice that the SA-CM cost function in this example has four global minima:  $f_0 = \pm(2 - j)$ , yielding  $y(n) = \pm s(n)$  and  $f_0 = \pm(-4 - 3j)$  yielding  $y(n) = \pm s(n - 1)$ .

### 4.2. Carrier Phase Offset Tolerance

Practical demodulators downconvert passband signals to near baseband signals using carrier recovery circuitry. However, this downconversion process is rarely exact and results

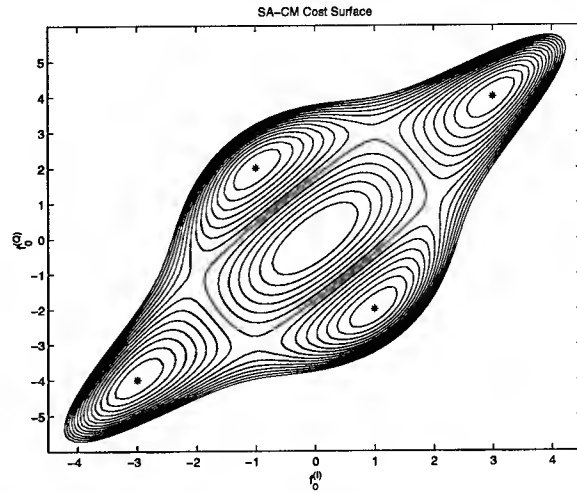


Figure 2: SA-CM cost function for example in Section 4.1.

in small carrier frequency and arbitrary phase offsets. We study the effect of static phase offsets on the MMSE performance of the three receivers in Figure 1.

For purposes of comparison, we hold the amount of hardware required to implement them constant. Since Receiver A is a real equalizer operating on real data, there is only one multiply required per tap. For Receiver B, we have a complex equalizer operating on complex data which will require four multiplies per tap. Receiver C uses a complex equalizer on real data, requiring two multiplies per tap. Hence, for a given level of hardware complexity, Receiver A allows the longest equalizer, and Receiver B the shortest. In the examples to follow, we choose filter lengths consistent with this constraint.

The MMSE performance of receiver A is a function of carrier phase error,  $\theta$ . This dependence can be seen in the example of Figure 3, where the channel impulse response coefficients are

$$\begin{aligned} &(-0.25 - 0.14j, \quad -0.3477 - 1.10j, \\ &\quad -0.41 + 0.31j, \quad -0.33 + 1.18j, \quad 0.09 + 1.17j) \end{aligned}$$

and 20dB SNR white Gaussian noise is considered. Receiver A outperforms receivers B and C for some phase offsets  $\theta$ , possibly due to its longer filter length. However, the MMSE performance of receivers B and C is independent of  $\theta$  since they employ equalizers with complex coefficients. As indicated above, Receiver C can have twice as many taps as receiver B for a given hardware complexity. Furthermore, as noted in [10], the MSE criterion of receiver C is less restrictive than that of receiver B, since only the real part of the mean squared recovery error is minimized. Hence, receiver C has the lowest MMSE performance of the three receivers for nearly all phases.

Note that the real-part extraction used in the SA-CM criterion relies on precise carrier frequency offset estimation. Unfortunately, the decoupling of equalization and carrier frequency recovery provided by CMA for complex sources [9] is lost in the SA-CM criterion. In this case, when carrier

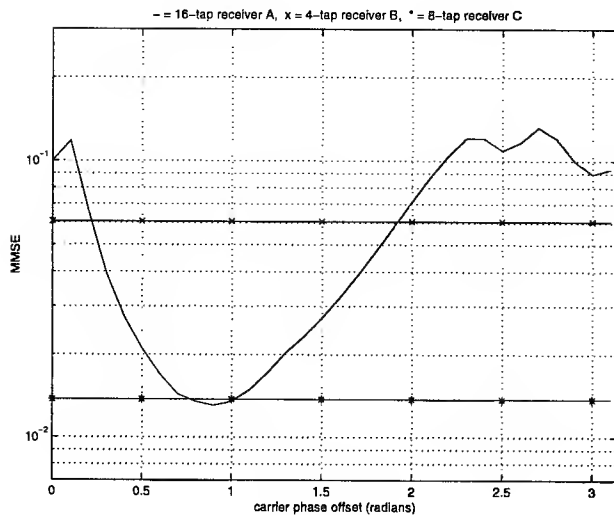


Figure 3: MMSE performance of receivers A, B and C with respect to carrier phase offset.

frequency recovery is imprecise, convergence of SA-CMA is not guaranteed.

#### 4.3. Vestigial Sideband Modulation

Vestigial Sideband Modulation (VSB) has a long history in analog communications and is particularly relevant to modern digital communications because the Advanced Television Systems Committee has adopted and upheld VSB as the modulation for high definition television (HDTV) in the United States. Reference [1] concludes that for VSB, a complex valued equalizer is unnecessary on the basis that only the in-phase carrier (I) is modulated with unique data. However, a real valued equalizer solution similar to receiver A will not take advantage of the channel phase diversity inherent in VSB.

One model for VSB modulation transmits real valued data through a complex valued pulse shaping filter where the quadrature component of the filter is roughly the hilbert transform of the in-phase component [3]. For example, a FIR model of the VSB pulse shaping filter is shown in Figure 4. Notice the hilbert transform impulse response in the quadrature axis and a single spike in the in-phase axis. The complex valued VSB pulse shaping filter applied to a real valued data source is a communication model amicable to receiver C in Figure 1. For our simulation comparison, we use a 2-VSB source with a combined channel-pulse shape filter whose frequency response is shown in Figure 5, with 40dB SNR additive Gaussian noise. Receivers A, B, and C are all baud-spaced, with 96 taps, 24 taps, and 48 taps, respectively. The length of the simulation is 50,000 iterations, with receivers A and B updated using CMA, while receiver C is updated using SA-CMA.

Figure 6 shows that the receiver C demonstrates the lowest MSE performance for this set of channel conditions. The resulting channel-equalizer response for receiver B, shown in Figure 7, does not result in a pure delay (see [7]). This

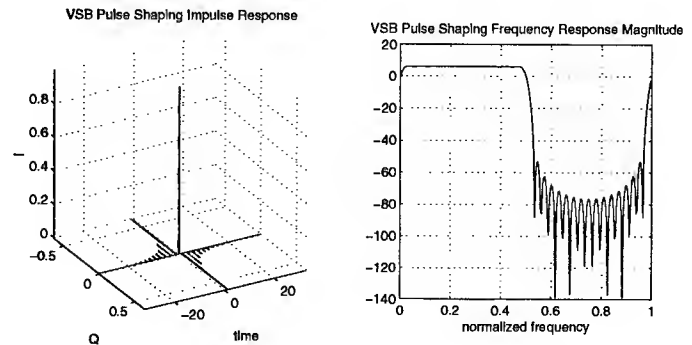


Figure 4: Impulse response and magnitude of frequency response of VSB pulse shaping filter.

explains why receiver B does not converge to an acceptable MSE performance. The channel-equalizer responses of receivers A and C do converge to near pure delays, with receiver C showing better MSE performance. However, for channel responses that require a longer equalizer span, a complex equalizer may be computationally prohibitive, and a real valued equalizer could be preferred.

#### 5. CONCLUSION

We have described a modification of the CM criterion for real valued sources applicable to receivers that employ complex valued signal processing. Our modification extracts the real part of the equalizer output and is thus called Single-Axis CM (SA-CM) criterion. We have shown that a finite length, baud-spaced equalizer minimizing the SA-CM criterion exploits the phase diversity in a complex valued channel and admits only desirable global minima under perfect equalizability conditions. Finally, we have provided simulations to demonstrate its feasibility for staggered modulation formats such as VSB signaling.

#### 6. FUTURE WORK

We have provided some motivation for the use of single-axis equalization for real valued and staggered modulation sources. Further application and study is warranted for more sophisticated equalizer architectures, such as IIR and decision feedback (DFE) equalizers. Carrier frequency offset is a practical concern in many equalizer designs. Thus, performance studies of single-axis equalization and SA-CMA under non-ideal carrier recovery is needed.

#### REFERENCES

- [1] Advanced Television Systems Committee Technology Group on Distribution, *Guide to the Use of the ATSC*

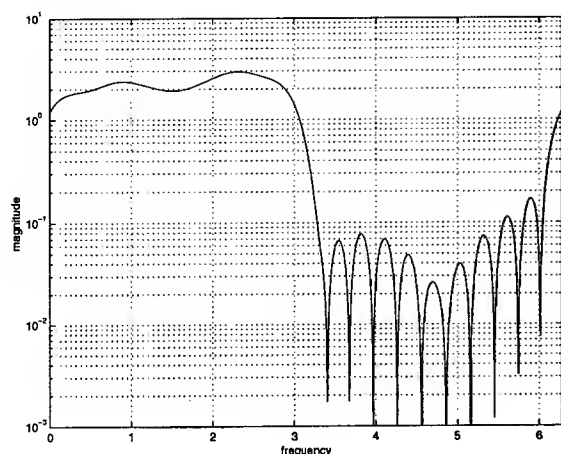


Figure 5: Frequency spectrum of combined channel-VSB pulse shape.

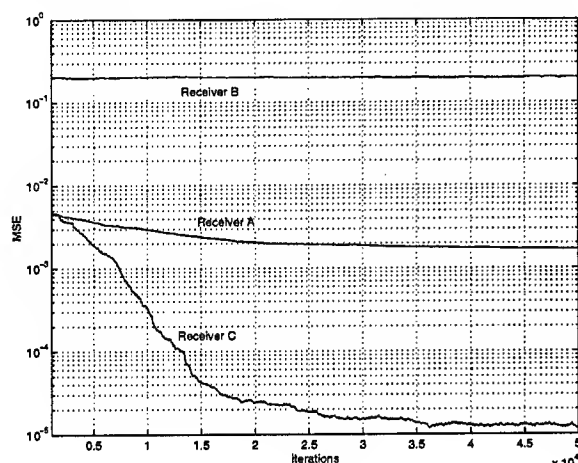


Figure 6: Transient MSE performance of receivers A, B, and C applied to VSB signaling.

*Digital Television Standard*, www.atsc.org, October 4, 1995.

- [2] I. Fijalkow, F. Lopez de Victoria, C.R. Johnson, Jr. "Adaptive fractionally-spaced blind equalization," *Proceedings of the IEEE Signal Processing Workshop*, Yosemite National Park, CA, pp.257-60, October 1994.
- [3] R.D. Gitlin, J.F. Hayes, S.B. Weinstein, *Data Communications Principles*, Plenum Press, New York, NY, 1992.
- [4] D. N. Godard, "Self-Recovering Equalization and Carrier Tracking in Two-Dimensional Data Communication Systems," *IEEE Transactions on Communications*, vol. 28, no. 11, pp.1867-1875, October 1980.
- [5] C.R. Johnson, Jr., P. Schniter, T.J. Endres, J.D. Behm, D.R. Brown, and R.A. Casas, "Blind equalization using the constant modulus criterion: a review,"

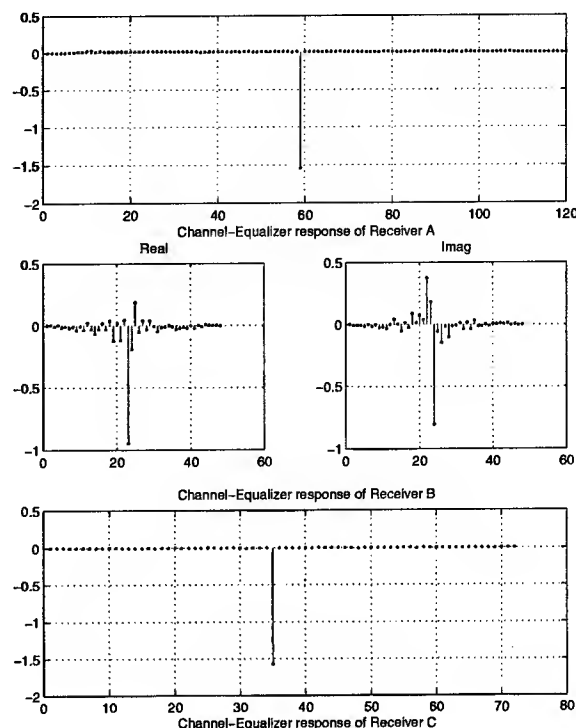


Figure 7: Combined channel-equalizer impulse response. Top: Receiver A. Middle: Receiver B. Bottom: Receiver C

*Proceedings of the IEEE*, vol.86, no. 10, pp.1927-1949, October 1998.

- [6] Y. Li, Z. Ding, "Global convergence of Fractionally Spaced Godard (CMA) Adaptive Equalizers," *IEEE Transactions on Signal Processing*, vol.44, no.4, pp.818-826, April 1996.
- [7] C. B. Papadias, "On the Existence of Undesirable Global Minima of Godard Equalizers," *Proceedings of the International Conference on Acoustics, Speech and Signal Processing*, Munich, Germany, pp.3941-3944, April 21-24, 1997.
- [8] L. Tong, G. Xu, T. Kailath, "Fast blind equalization via antenna arrays," *Proceedings of the International Conference on Acoustics, Speech and Signal Processing*, Minneapolis, MN, vol. 4, pp.272-275, April 1993.
- [9] J. R. Treichler, M. G. Larimore, J. C. Harp, "Practical Blind Demodulators for High-Order QAM Signals," *Proceedings of the IEEE*, vol.86, no. 10, pp.1907-1926, October, 1998.
- [10] J. C. Tu, "Optimum MMSE Equalization for Staggered Modulation," *Record of Asilomar Conference on Signals, Systems and Computers*, Pacific Grove, CA, pp.1401-1406, November 1-3, 1993.

# A NOVEL MODULATION METHOD FOR SECURE DIGITAL COMMUNICATIONS

Arnt-Børre Salberg and Alfred Hanssen

University of Tromsø, Physics Department  
Electrical Engineering Group, N-9037 Tromsø, Norway

## ABSTRACT

We present a new digital modulation technique that introduces covertness in digital communications. The basic principle is to transmit realizations of a stochastic process in such a manner that the transmitted waveform appears noise-like. In this paper, we have chosen to express the transmitted waveform in a subspace formalism. This allows for an elegant geometrical interpretation of the waveform, and it naturally suggests a simple and accurate matched subspace detector for the receiver. The technique is demonstrated by numerical simulations, and a comparison with an optimal Neyman-Pearson detector shows that our simple subspace detector yields a high-quality and reliable receiver for the modulated signal.

## 1. INTRODUCTION

An obvious way of introducing covertness in digital communications, is to ensure that the transmitted waveform appears noise-like. Spread-spectrum techniques e.g. [1], apply a known quasi-stochastic spreading sequence to obtain some degree of privacy. To decode the signal, the receiver must have complete knowledge about the spreading sequence, and it must be strictly synchronous with the transmitter. In addition, a simple spectrogram analysis may detect the pulses and disclose the existence of the transmission.

Salberg and Hanssen in [2] proposed the following low-probability-of-intercept method for encoding digital information. Transmit a realization of a stochastic process  $X_0(t)$ ,  $0 \leq t < T$  to represent bit zero, and a realization of another stochastic process  $X_1(t)$ ,  $0 \leq t < T$  to represent bit one. Here  $T$  is the symbol duration. Thus, rather than altering aspects of a deterministic carrier signal, realizations of two different stochastic processes are transmitted. This has the effect that two subsequent equal source bits have different transmitted waveforms. In addition, two different source bits have similar waveforms, due to the fact that they are close in a statistical sense. The transmitted waveform representing a bit string will thus appear noise-like, and it contains no repetitions or periodicities. Moreover, the waveform contains no discontinuities, so the pulse length is also

hidden. Since the transmitted baseband waveform is noise-like, a transmission would not attract the attention of unfriendly receivers. It is obvious that this signaling method adds an extra (physical) layer of security in digital communication, thus reducing the risk of eavesdropping.

In this paper we will generalize the technique suggested in [2]. We have chosen to express the waveform generator by means of an orthonormal basis. The background stochastic sequences are generated by a redundant linear transformation of a stochastic coefficient vector, and a transmitted "pulse" is simply the stochastic sequence expressed in the chosen basis. The benefit of such an encoding is that very simple and efficient decoders can be constructed by means of subspace projections onto the two different subspaces spanned by the basis waveforms.

## 2. STOCHASTIC PROCESS SHIFT KEYING

Let  $\{a_n\}_{n=-\infty}^{\infty}$  be the source bit sequence. The transmitted waveform for an infinite duration *Stochastic Process Shift Keying* signal suggested in [2] can then be written as

$$X(t) = \sum_{n=-\infty}^{\infty} [a_n X_0(t) + (1 - a_n) X_1(t)] u(t - nT - \lambda) \quad (1)$$

where  $u(t)$  is a unit amplitude rectangular pulse of duration  $T$ . Here we model the source as a wide-sense stationary stochastic sequence with 0 and 1 as possible outcomes, mean value  $E[a_n] = \mu_a$ , correlation sequence  $E[a_n a_{n+k}] = R_a(k)$ , and  $\lambda$  is a uniformly distributed random variable  $\lambda \sim \mathcal{U}[0, T]$  independent of  $a_n$ .

A more general strategy is to write the transmitted pulse waveform as a linear combination of some basis waveforms  $f_k(t) \in \mathcal{F}$ ,  $k = 1, 2, \dots$ , where  $\mathcal{F}$  is a function space that satisfies some desired properties. Assume that to transmit bit zero we use waveforms from a subspace  $\mathcal{G}_0 \subset \mathcal{F}$ , and to transmit bit one we use waveforms from a subspace  $\mathcal{G}_1 \subset \mathcal{F}$ . The subspaces  $\mathcal{G}_0$  and  $\mathcal{G}_1$  have rank  $M_0$  and  $M_1$ , respectively, and in general  $\mathcal{G}_0 \cap \mathcal{G}_1 \neq \emptyset$  which means that bit zero and bit one may have common basis waveforms. Let the number of basis waveforms in the set  $\mathcal{G}_0 \cup \mathcal{G}_1$  be

$2M$ . Thus the transmitted pulse is

$$s_i(t) = \mathbf{f}^T(t) \mathbf{G}_i \mathbf{s}_i, \quad i = 0, 1 \quad (2)$$

where  $\mathbf{f}(t) = [f_1(t), f_2(t), \dots, f_{2M}(t)]^T$  is a vector consisting of the basis waveforms,  $\mathbf{s}_i = [s_{i,1}, s_{i,2}, \dots, s_{i,M_i}]^T$  is a random vector drawn from a multivariate probability density  $p_{\theta_i}(\mathbf{s})$ , where  $\theta_i$  is a relevant parameter vector, and  $\mathbf{G}_i = [\mathbf{g}_1^{(i)}, \mathbf{g}_2^{(i)}, \dots, \mathbf{g}_{M_i}^{(i)}]$  is a  $2M \times M_i$  matrix of rank  $M_i$ . The transmitted waveform  $X(t)$  can then be written as

$$X(t) = \sum_{n=-\infty}^{\infty} \mathbf{f}^T(t - nT - \lambda) \cdot [a_n \mathbf{G}_0 \mathbf{s}_{0,n} + (1 - a_n) \mathbf{G}_1 \mathbf{s}_{1,n}], \quad (3)$$

where  $\mathbf{s}_{i,n}$  is the stochastic parameter vector at time  $n$ .

### 3. BASIS FUNCTIONS

In our attempt to choose a basis function space  $\mathcal{F}$  there are several aspects that we must consider. For instance, we often want the basis functions  $f_k(t)$  to be compactly supported such that a transmitted pulse is time limited. The pulse length of the basis waveforms does not have to be equal to the symbol duration  $T$ , which means that a transmitted pulse can overlap with neighboring pulses. Furthermore, the decoding can be made simple if the basis waveforms in  $\mathcal{F}$  are orthogonal, i.e.

$$\int_{T_s} f_k(t) f_j(t) dt = \alpha \delta_{k,j}, \quad (4)$$

where  $\alpha$  is a constant,  $\delta_{k,j}$  is Kronecker's delta, and  $T_s$  is the pulse length. As an aspect of low-probability-of-intercept we require that the waveforms are chosen such that the transmitted baseband signal is noise-like, and if the waveforms do not contain any discontinuities that can compromise the pulse length we have an additional security at the waveform level.

Yi and Powers in [3] proposed a wavelet-based orthogonal modulation code set where the code set consists of various orthogonal scaling functions and mother wavelets.

#### 3.1. Orthogonal Modulation Code

Yi and Powers [3] used the Hadamard matrices to design orthogonal code sets. The Hadamard matrices are defined as

$$H_2 = \begin{bmatrix} 1 & 1 \\ 1 & -1 \end{bmatrix}, \quad H_{2n} = \begin{bmatrix} H_n & H_n \\ H_n & -H_n \end{bmatrix} \quad (5)$$

where  $n$  is an integer and the dimensions of  $H_n$  are  $2n \times 2n$ . Since the row vectors of the Hadamard matrix are orthogonal, the Hadamard matrix yields an efficient tool to construct orthogonal basis waveforms.

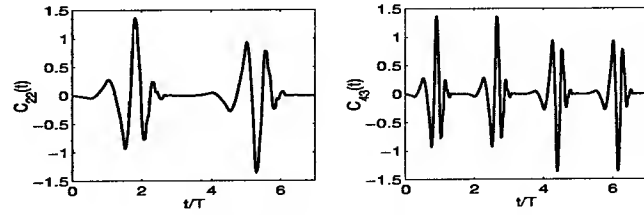


Figure 1: Example of mutually orthogonal basis waveforms  $C_{22}(t)$  and  $C_{43}(t)$  using Daubechies 4 wavelets.

In the discrete wavelet transform the scaled and translated orthogonal dyadic wavelet  $\psi_{j,k}(t)$  is defined as

$$\psi_{j,k}(t) = 2^{-j/2} \psi(2^{-j}t - k) \quad (6)$$

where  $\psi(t)$  is the mother wavelet,  $j$  is a scale index, and  $k$  is a translation index. For integers  $j, k, m$  and  $n$  we have that the inner product obeys

$$\langle \psi_{j,k}(t), \psi_{m,n}(t) \rangle = \delta_{j,m} \delta_{k,n}. \quad (7)$$

The scaling functions  $\phi_{j,k}(t)$  are orthogonal only across translation but not across scale,

$$\langle \phi_{j,k}(t), \phi_{j,n}(t) \rangle = \delta_{k,n}. \quad (8)$$

At specific scales and several translations the wavelets are orthogonal to scaled and translated scalings functions. For any  $j \leq m$ , we have

$$\langle \psi_{j,k}(t), \phi_{m,n}(t) \rangle = 0. \quad (9)$$

From the properties discussed above, Yi and Powers [3] proposed the following wavelet-based orthogonal modulation code

$$\begin{aligned} C_{1,1}(t) &= A \phi_{j,0}(t) \\ C_{1,2}(t) &= A \psi_{j,0}(t) \\ C_{m,n}(t) &= A \sum_{k=0}^{m-1} H_m(n, k+1) \psi(2^{-(j-p)}t - kl) \end{aligned} \quad (10)$$

where  $A$  is a constant,  $m$  is a constant which must be a power of 2,  $p = \log_2 m$ ,  $n$  is  $1 \leq n \leq m$ , and  $l$  is the time domain support length of the wavelet  $\psi_{j,0}(t)$  (which equals the pulse length  $T_s$ ). Thus,  $2M$  orthogonal wavelet basis waveforms  $C_{1,1}(t), C_{1,2}(t), \dots, C_{M,M}(t)$  span  $\mathcal{G}_0 \cup \mathcal{G}_1$ , and we select at least  $M_i$  of these orthogonal basis waveforms to span  $\mathcal{G}_i$ . The selection of basis waveforms is performed by the matrices  $\mathbf{G}_0$  and  $\mathbf{G}_1$ .

Fig. 1 shows examples of orthogonal basis waveforms based on Daubechies 4 wavelets and scaling function [4]. From the example we see that the waveforms become "sharper" as  $M$  increases, and that they thus contain higher frequency components. For Daubechies 4 wavelets, we have that  $T_s = lT = 7T$ . Thus, the information carrying pulses will overlap substantially in time.

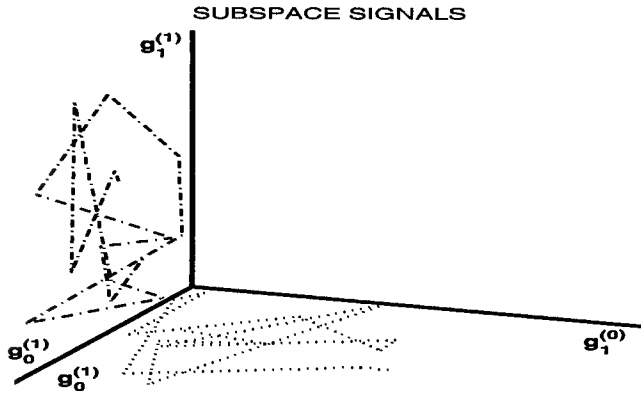


Figure 2: Trajectories of the subspace signals  $\mathbf{x}_0$  (dotted) and  $\mathbf{x}_1$  (dash-dotted).

#### 4. DECODING

The vector representation of the transmitted pulse is

$$\mathbf{x}_i = \sum_{k=1}^{M_i} s_{i,k} \mathbf{g}_k^{(i)} = \mathbf{G}_i \mathbf{s}_i, \quad i = 0, 1. \quad (11)$$

In this case, the signal  $\mathbf{x}_i$  is known to lie in the  $M_i$  dimensional linear subspace  $\langle \mathbf{G}_i \rangle$  spanned by the columns of  $\mathbf{G}_i$ . This is illustrated in Fig. 2 where the dotted line is the trajectory of the subspace signal  $\mathbf{x}_0$ , the dash-dotted line the is the trajectory of the subspace signal  $\mathbf{x}_1$ ,  $M_0 = M_1 = 2$  and  $2M = 3$ . From the figure we see the randomness of the signals  $\mathbf{x}_0$  and  $\mathbf{x}_1$ , and that  $\mathbf{x}_i$  is in the subspace spanned by the columns of  $\mathbf{G}_i = [\mathbf{g}_0^{(i)} \mathbf{g}_1^{(i)}]$ . The matrix  $\mathbf{G}_i$  can be chosen to introduce redundancy in the transmitted symbol  $\mathbf{x}_i$ , and the elements in  $\mathbf{x}_i$  are then linear combinations of the elements in  $\mathbf{s}_i$ . We see that the  $\mathbf{g}_k^{(i)}$  direction is weighted by  $s_{i,k}$ , and we now have a correspondence between the physical time-domain and a  $2M$ -dimensional signal space.

We define the projection operator as [5]

$$\mathbf{P}_{G_i} = \mathbf{G}_i (\mathbf{G}_i^T \mathbf{G}_i)^{-1} \mathbf{G}_i^T, \quad i = 0, 1 \quad (12)$$

so that  $\mathbf{P}_{G_i} \mathbf{r}$  is a projection of the vector  $\mathbf{r}$  onto the subspace  $\langle \mathbf{G}_i \rangle$ . If the subspaces  $\langle \mathbf{G}_0 \rangle$  and  $\langle \mathbf{G}_1 \rangle$  are disjoint, the columns of  $\mathbf{G}_0$  and  $\mathbf{G}_1$  are linearly independent. A stronger condition is orthogonality, which means that  $\mathbf{G}_0^T \mathbf{G}_1 = \mathbf{0}$ . For orthogonal subspaces we have

$$\mathbf{P}_{G_0} \mathbf{G}_0 = \mathbf{G}_0 \quad \text{and} \quad \mathbf{P}_{G_0} \mathbf{G}_1 = \mathbf{0} \quad (13)$$

$$\mathbf{P}_{G_1} \mathbf{G}_1 = \mathbf{G}_1 \quad \text{and} \quad \mathbf{P}_{G_1} \mathbf{G}_0 = \mathbf{0} \quad (14)$$

and we see that an orthogonal projection has a null space that is orthogonal to its range.

#### 4.1. Detection

Given a transmitted time-domain waveform  $s_i(t)$ , we assume that this signal is contaminated by an additive disturbance  $n(t)$ , so that the waveform at the receiver input is  $r(t) = s_i(t) + n(t)$ , where  $n(t)$  is a zero-mean, Gaussian white noise process with power spectral density  $S_n(\omega) = \mathcal{N}_0/2$ ,  $\forall \omega$ . The vector representation of the signal plus noise is  $\mathbf{r} = \mathbf{x}_i + \mathbf{n}$ , where the elements of the received vector  $\mathbf{r}$  are

$$r_j = \int_{T_s} r(t) f_j(t) dt = x_{i,j} + n_j, \quad j = 1, \dots, 2M. \quad (15)$$

Since the noise is a zero-mean Gaussian process,  $n_j$  is also Gaussian with  $E\{n_j\} = 0$  and  $\text{Var}\{n_j\} = \mathcal{N}_0/2$ .

Assume that the stochastic coefficient vector  $\mathbf{s}_i$  is multivariate Gaussian  $N[\mathbf{m}_{s,i}, \mathbf{R}_{s,i}]$  and that the noise vector  $\mathbf{n}$  is distributed as  $N[\mathbf{0}, (\mathcal{N}_0/2)\mathbf{I}]$ . Define  $E[\mathbf{x}_i] = \mathbf{m}_i = \mathbf{G}_i \mathbf{m}_{s,i}$  and  $E[(\mathbf{x}_i - \mathbf{m}_i)(\mathbf{x}_i - \mathbf{m}_i)^T] = \mathbf{R}_{x,i} = \mathbf{G}_i \mathbf{R}_{s,i} \mathbf{G}_i^T$ , then if  $s_i(t)$  is sent we have that the received vector  $\mathbf{r}$  is distributed as  $N[\mathbf{m}_i, \mathbf{R}_{x,i} + (\mathcal{N}_0/2)\mathbf{I}]$ . Note that the matrix  $\mathbf{R}_{x,i}$  may be singular, depending on  $\mathbf{G}_i$ . In that case a bias term  $\lambda \mathbf{I}$ , where  $\lambda \ll 1$ , must be added to regularize  $\mathbf{R}_{x,i}$ .

In general, a Neyman-Pearson hypothesis test (e.g., [5]) may be applied in the decoding, since we assume that all relevant probability densities are known to the intended receiver.

Another, but suboptimum, detector that may be used is the so called *matched subspace detector* (MSD) [5]. Scharf proposed the MSD to detect an unknown deterministic subspace signal in a known subspace. We have extended the MSD to classify *stochastic subspace signals* in known subspaces. The basic idea is to regard a stochastic subspace signal as an unknown deterministic subspace signal. In case of two different classes, the extended MSD will project the received vector  $\mathbf{r}$  onto the subspaces  $\langle \mathbf{G}_0 \rangle$  and  $\langle \mathbf{G}_1 \rangle$ . The statistic  $\mathbf{r}^T \mathbf{P}_{G_i} \mathbf{r}$  is clearly a maximal invariant statistic [5], and the decision criterion is that we choose class  $\Omega_0$  if

$$\mathbf{r}^T \mathbf{P}_{G_0} \mathbf{r} > \mathbf{r}^T \mathbf{P}_{G_1} \mathbf{r}, \quad (16)$$

and otherwise choose class  $\Omega_1$ . The detector measures the amount of the received energy that resides in subspace  $\langle \mathbf{G}_i \rangle$ , and then chooses the class corresponding to the subspace containing the largest amount of energy. The benefit of such a detector compared with the Neyman-Pearson detector is that the decision criterion is independent of the additive noise variance ( $\mathcal{N}_0/2$ ). Obviously, the particular choice of subspace matrices  $\mathbf{G}_0$  and  $\mathbf{G}_1$  will influence on the performance of the system.

An analytic expression for the bit-error probability (BEP) of the Neyman-Pearson detector for Gaussian signal and noise has been given in [6]. The expression is on integral form, and may thus be evaluated numerically.



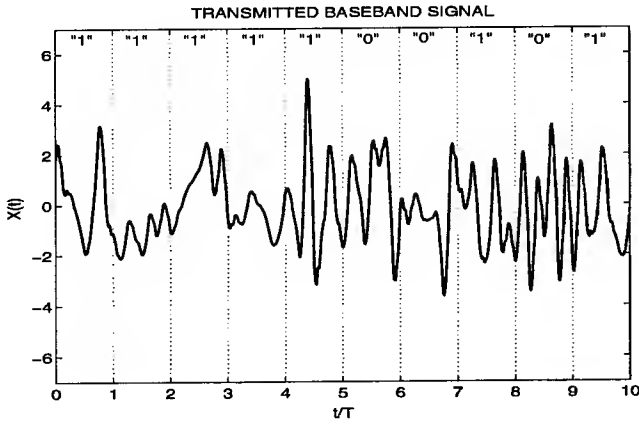


Figure 3: Message '1111100101' encoded by means of orthonormal basis waveforms constructed in Eq. (10) with  $M = 4$ .

## 5. SIMULATIONS

To demonstrate the proposed digital modulation technique, we now present some numerical simulations.

In our numerical simulations we use orthonormal wavelet basis functions given by Eq. (10). The basis functions are constructed by Daubechies 4 wavelets, which yields a time domain support length  $l = 7$ , and the scale index was  $j = 0$ . The subspaces matrices  $\mathbf{G}_0$  and  $\mathbf{G}_1$  are orthogonal to each other and have orthonormal columns. Furthermore,  $p_{\theta_0}(\mathbf{s}) = p_{\theta_1}(\mathbf{s})$  is multivariate Gaussian  $N[\mathbf{0}, \mathbf{R}_s]$ , which yields  $E\{\mathbf{s}_0^T \mathbf{s}_0\} = E\{\mathbf{s}_1^T \mathbf{s}_1\}$ . The random vector  $\mathbf{s}_0$  and  $\mathbf{s}_1$  are generated as realizations of an AR(2)-process with parameters  $\theta_0 = \theta_1 = [a_1, a_2, \sigma^2]^T = [0.1, 0.35, 1]^T$ . The matrices  $\mathbf{G}_0$  and  $\mathbf{G}_1$  are constructed from the orthonormal eigenvectors of the  $2M \times 2M$  covariance matrix of an AR(2) process with  $a_1 = 0.81, a_2 = 0.35$  and  $\sigma^2 = 1$ . This is a simple way of constructing the subspace matrices, but obviously not the only possibility.

Fig. 3 shows an example of the transmitted waveform for the message '1111100101'. The encoder applies orthonormal basis waveforms constructed from Eq. (10) with  $M = 4$ , which yields 8 different basis waveforms. Observe that two subsequent equal source bits have different waveforms, since the basis waveforms are weighted by a stochastic vector  $\mathbf{x}_i$ . Note also that the pulse length is hidden, and that there are no periodicities in the information carrying signal.

The average pulse energy is

$$E_b = E\{\langle s_i(t), s_i(t) \rangle\} = E\left\{\int_{T_s} |s_i(t)|^2 dt\right\}. \quad (17)$$

Using Eq. (2) and (11) we have that

$$E_b = E\left\{\int_{T_s} \mathbf{x}_i^T \mathbf{F}(t) \mathbf{x}_i dt\right\} \quad (18)$$

where

$$\mathbf{F}(t) = \begin{bmatrix} f_1(t)f_1(t) & \dots & f_1(t)f_{2M}(t) \\ \vdots & & \vdots \\ f_{2M}(t)f_1(t) & \dots & f_{2M}(t)f_{2M}(t) \end{bmatrix}. \quad (19)$$

Furthermore, using the orthonormality property of the basis functions and that  $\mathbf{G}_i^T \mathbf{G}_i = \mathbf{I}$  we find that

$$E_b = E\{\mathbf{x}_i^T \mathbf{x}_i\} = E\{\mathbf{s}_i^T \mathbf{G}_i^T \mathbf{G}_i \mathbf{s}_i\} = E\{\mathbf{s}_i^T \mathbf{s}_i\}. \quad (20)$$

Thus, as in conventional communications we may define the signal-to-noise ratio (SNR) as  $\text{SNR} = E_b/\mathcal{N}_0$ , which in our case can be written as  $\text{SNR} = \text{tr}\{\mathbf{R}_s\}/\mathcal{N}_0$ , where  $\text{tr}\{\cdot\}$  denotes the trace.

Fig. 4 shows the exact BEP of the Neyman-Pearson detector with all parameters known (full lines), and Monte Carlo simulated BEP (20000 repetitions) of the extended MSD (crosses). In curve (i) 8 orthonormal basis waveforms are used, and  $\mathbf{G}_i$  has dimension  $8 \times 4$ . Curve (ii) shows the BEP with 40 orthonormal basis waveforms, and  $\mathbf{G}_i$  of dimension  $40 \times 20$ . From Fig. 4 we see, as expected, that the BEP decreases as a function of increasing SNR. Furthermore, notice that the suboptimal extended MSD has a performance close to that of the optimal Neyman-Pearson detector. This is a remarkable result since the Neyman-Pearson detector is assumed to have knowledge about the measurement noise  $\mathcal{N}_0/2$  and the covariance matrix  $\mathbf{R}_s$ . The reason for this close-to-optimal performance of the extended MSD is the orthogonality of the subspaces  $\langle \mathbf{G}_0 \rangle$  and  $\langle \mathbf{G}_1 \rangle$ . Since the subspace matrices encode the random vector  $\mathbf{s}$ , the extended MSD has all the relevant information available. The lack of knowledge of the noise variance  $\mathcal{N}_0/2$  does not influence on the performance of the extended MSD, since the decision is based on the energy of the received vector  $\mathbf{r}$  in each subspace. If we however choose  $\theta_0 \neq \theta_1$  then the Neyman-Pearson detector will outperform the extended MSD for any choice of subspaces  $\langle \mathbf{G}_0 \rangle$  and  $\langle \mathbf{G}_1 \rangle$ .

For low-probability-of-intercept communications, the dimension of  $\mathbf{G}_i$  must be chosen such that the signal power per unit bandwidth is below the noise spectral density. In the simulations above,  $M = 20$  implies the use of waveforms with higher frequency components than for  $M = 4$ . Thus, larger values of  $M$  spreads the transmitted signal over a wider frequency band. Since the energy of the transmitted baseband waveforms in case of  $M = 4$  and  $M = 20$  are equal, the signal power per unit bandwidth is lower for the case of  $M = 20$ .

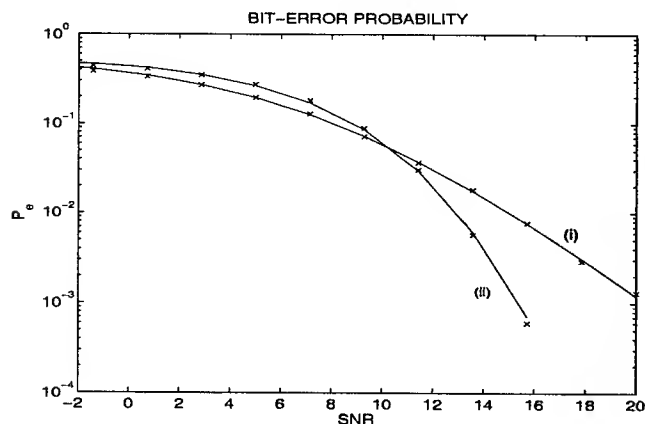


Figure 4: Bit-error probability as a function of SNR, (i)  $M = 4$  and (ii)  $M = 20$ . Crosses are Monte Carlo simulations of the extended MSD, and full curves are exact theoretical results of the Neyman-Pearson detector.

Fig. 5 shows a spectrogram (in dB) of the transmitted baseband signal in Fig. 3. The horizontal axis is the normalized time axis, and the vertical axis is the frequency axis, normalized with respect to the maximum frequency  $f_{max}$  of  $X(t)$ . From the spectrogram we clearly see that there is no structures that can disclose the transmitted bit sequence, nor disclose that an information carrying signal is actually being transmitted.

## 6. CONCLUSIONS

We have presented a new digital modulation technique that offers some degree of security at the waveform level. The transmitted waveform is noiselike, and would therefore not attract the attention of unfriendly receivers. It is demonstrated that simple and efficient detectors can be constructed by means of a subspace formalism, and that in some cases the performance of these suboptimal detectors equals that of the optimum Neyman-Pearson detector. The extension of the proposed technique to multi-user communication is straightforward.

## REFERENCES

- [1] J. G. Proakis, *Digital communication*, McGraw-Hill, New York, 1995.
- [2] A. B. Salberg and A. Hanssen, "Secure digital communications by means of stochastic process shift keying," in *Proc. 33rd Asilomar Conf. Signals, Syst., Comp.*, Pacific Grove, CA, 1999.

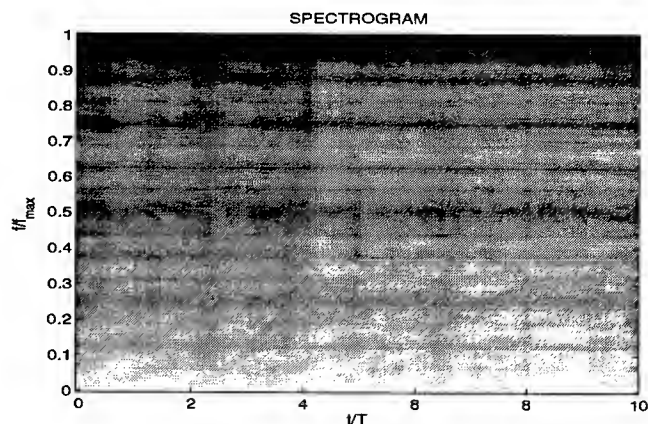


Figure 5: Spectrogram of the transmitted baseband signal in Fig. 3.

- [3] E. J. Yi and E. J. Powers, "Wavelet-based orthogonal modulation code," in *Proc. 33rd Asilomar Conf. Signals, Syst., Comp.*, Pacific Grove, CA, 1999, pp. 1632–1636.
- [4] S. Mallat, *A Wavelet Tour of Signal Processing*, Academic, San Diego, CA, 1998.
- [5] L. L. Scharf, *Statistical Signal Processing: Detection, Estimation, and Time Series Analysis*, Addison-Wesley, Reading, Mass., 1991.
- [6] K. Fukunaga and T. F. Krile, "Calculation of Bayes' recognition error for two multivariate Gaussian distribution," *IEEE Trans. Comp.*, vol. C-18, pp. 220–229, Mar. 1969.

# A MULTITIME-FREQUENCY APPROACH FOR DETECTION AND CLASSIFICATION OF NOISY FREQUENCY MODULATIONS

M. COLAS<sup>1</sup>, G. GELLE<sup>1</sup>, J. GALY<sup>2</sup>, G. DELAUNAY<sup>1</sup>

<sup>1</sup> L.A.M-URCA  
BP 1039  
51687 Reims cedex 2, France

e-mail : [guillaume.gelle@univ-reims.fr](mailto:guillaume.gelle@univ-reims.fr)

<sup>2</sup> L.I.R.M.M. UMR CNRS 5506  
165 rue Ada  
34392 Montpellier cedex 05, France

e-mail : [galy@lirmm.fr](mailto:galy@lirmm.fr)

## ABSTRACT

In this communication, Time Varying Higher Order Spectra and specifically multitime-frequency representations have been used for detection and classification purpose. A new detector is presented for frequency modulations disrupted by multiplicative and additive noise. Statistical study is performed and corresponding simulations are presented. An extension to multiple hypothesis testing is also presented to classify neighboring frequency modulations in a context of multiplicative and additive noise. Some simulations illustrate the performances of our approach comparing to the similar second order approach.

## 1. INTRODUCTION

Time-Frequency and Higher Order Spectra have been intensively studied during these last few years. The first involves time varying signals and depicts the evolution of the power spectral density through the time. However, second order statistics do not take into account the non-linear phenomena and perfectly describes only the linear systems and the gaussian processes. For non-gaussian signals and non-linear systems analysis, many techniques based on HOS were reported. Usually, this method required stationary assumption. Recently, Time Varying Higher Order Spectra (TVHOS) are defined [1],[2] and permit to analyze non-linear time varying signals. In this paper, we present two new detection / classification algorithms based on TVHOS and applied to frequency modulations disrupted by a real multiplicative noise in an additive complex noise.

## 2. TIME VARYING HIGHER ORDER SPECTRA

Many definitions of TVHOS can be found in the literature, they differ in particular in the lag separation between the time or frequency terms used for product. They can also differ in the number of conjugated terms and with the used space of representation as well : time-multifrequency space or multitime-frequency space. The user's aim will lead him to decide upon the type of representation whether he chooses to set out the non linear phenomena or to preserve the time-frequency accuracy; for example in modulation cases. To reduce the computational cost, it is customary to consider only a slice of TVHOS. Sliced

TVHOS (STVHOS) were first introduced by Fonollosa and Nikias in [3] and were defined as particular slices of the Wigner-Multispectrum. In practice, the principal slice of the Wigner-Trispectrum is the one most used for signal analysis because it is a real representation that contains all the autoterms of the signal.

A computationally efficient implementation is given in the frequency domain by :

$$SWD4_{x(n)}(n, \gamma) = \int X^2\left(\gamma + \frac{\gamma_1}{4}\right) X^{*2}\left(\gamma - \frac{\gamma_1}{4}\right) e^{-j\pi\gamma_1 n} d\gamma_1 \quad (1)$$

Simultaneously, Stankovic [4] proposed a multitime-frequency definition of Wigner Higher Order Distributions as for the fourth order :

$$MTWD4_{x(n)}(\underline{n}, \gamma) = \sum_k x^*(n1 + n2 + n3 + k) x(n3 - k) \\ x(n2 - k) x^*(-n1 + k) e^{-j8\pi\gamma k}$$

For computational purpose, evaluation of the *MTWD4* can be done by considering only the principal temporal slice given by  $n = -n1 = n2 = n3$ . Hence, we obtain the L-Wigner Ville distribution :

$$LW4_{x(n)}(n, \gamma) = \sum_k x^2(n+k) x^{*2}(n-k) e^{-j8\pi\gamma k} \quad (2)$$

which is a dual formulation of (1). Due to its good localization properties, even for non-linear frequency modulation, *LW4* has been extensively studied by Boashash in [5] for deterministic time varying signal processing. When dealing with random non stationary signals, Boashash defined the Moment and the Cumulant Wigner-Trispectrum (*MWD4* and *CWD4*) as :

$$MWD4_{(n)}(n, \gamma) = \sum_k m_4(n, k) e^{-j8\pi\gamma k} \\ CWD4_{(n)}(n, \gamma) = \sum_k c_4(n, k) e^{-j8\pi\gamma k} \quad (3a \& b)$$

with  $m_4(n, k) = E\{x^2(n+k) x^{*2}(n-k)\}$  ( $E$  denotes the statistical expectation) and  $c_4(n, k) = Cum\{x^2(n+k) x^{*2}(n-k)\}$  ( $Cum$  represents the cumulant operator). [5] shows that this 2

formulations are helpful in multiplicative noise signals for instantaneous frequency law estimations.

#### Signal analysis with TVHOS4

Let us consider the following model of signal :

$$x(n) = b_m(n).e^{j(\Phi(n)+\Theta)} \quad (4)$$

where  $b_m(n)$  is a zero mean white gaussian band limited process with variance  $\sigma_{bm}^2$ .  $\Theta$  is a random phase uniformly distributed in  $[0, 2\pi]$ .  $\Phi(n)$  is a polynomial function of the time index.

This model of signal have received an increasing interest these last few years, in particular in applications such as Radar and Sonar where, in addition to the Doppler effect, the returned signal is subjected to amplitude modulation caused by the changing orientation of non-point target [6]. Moreover, polynomial phase signal disrupted by a multiplicative noise provide an efficient model for speech analysis due to time varying amplitude produced by speech resonance [7]. Finally, acoustical analysis of transient signals produced by mechanical systems as been shown to follow such a model [8].

If we calculate the Wigner distribution of the signal (4), we obtain  $WS_{x(n)}(n, \gamma) = \sigma_{bm}^4$ . So, Wigner Spectrum is unable to characterize such signal. To perform the analysis of this model of signal and generally for non linear, non stationary signals, higher order approaches become necessary and the results obtained using the signal (4) are  $MWD4_{x(n)}(n, \gamma) = \sigma_{bm}^4 (LW4_{\Phi(n)}(n, \gamma) + 2)$  for a real white gaussian noise process  $b_m$ .

Multitime-frequency representations, using the spectral dependencies between negative and positive frequencies allow, for real random signals, the construction of a non oscillatory interference localized on the Instantaneous Frequency Law (IFL)  $\Phi'(n)$  as shown on figure 1.

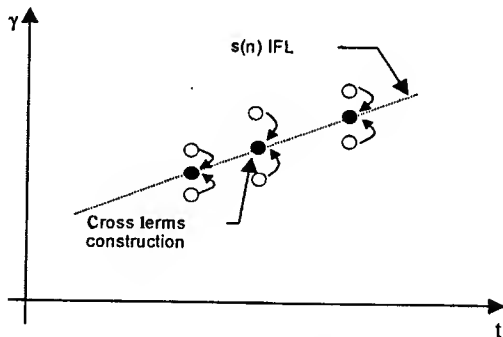


Figure 1 : interferences geometry in real multiplicative noise

Due to this property, we can conclude that  $MWD4$  gives a better estimate frequency law  $\Phi'(n)$  of the signal (4). Similar conclusion can be made for  $CWD4$ .

We illustrate this interferences property of TVHOS for the following signal :

$$x_2(n) = b_m(n)s(n) \quad (5)$$

with  $s(n) = e^{j2\pi(-a.\cos(2\pi.f_0.n)+f_1.n)+\phi}$ . For this signal and for a band limited real gaussian multiplicative noise with PSD  $BM(\gamma)$ , theoretical  $MWD4$  is given by :

$$MWD4_{s(n)}(t, \gamma) = \sigma_m^4 LW4_{s(n)}(t, \gamma) + LW4_{s(n)}(t, \gamma)^* BM(2\gamma)^* BM(2\gamma)$$

whereas theoretical  $WS$  is  $WS_{s(n)}(t, \gamma) = WVD_{s(n)}(t, \gamma)^* BM(\gamma)$

Figures 2,3,4 show the highly resolution of the fourth order time-frequency representation due to the previously mentioned property.

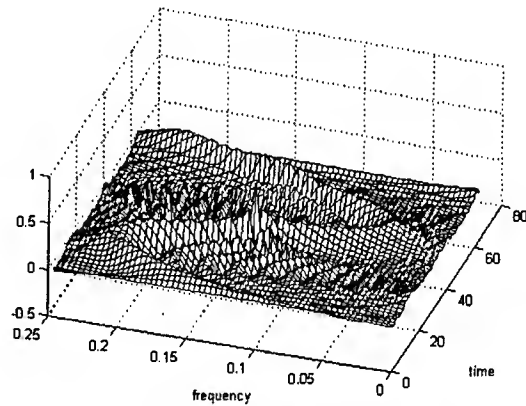


Figure 2 :  $MWD4$  for signal  $x_2$

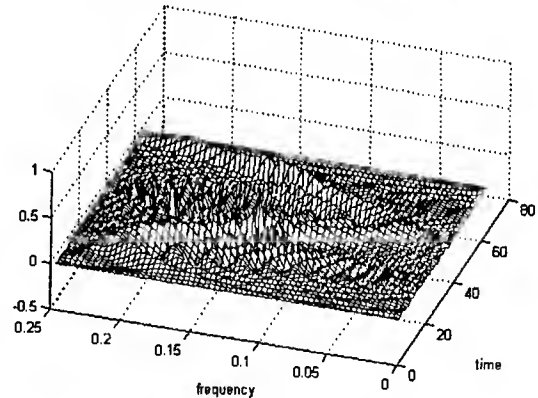


Figure 3 :  $CWD4$  for signal  $x_2$

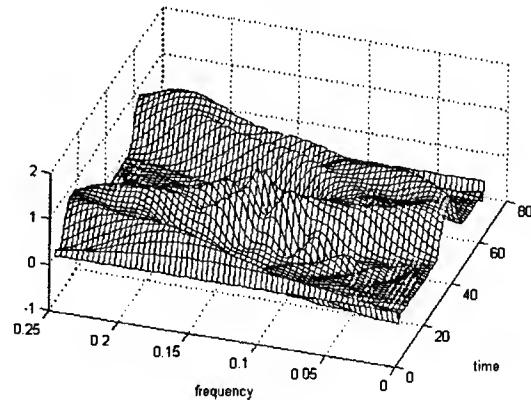


Figure 4 :  $WS$  for signal  $x_2$

Considering these results, it seems that the good localization properties of Multitime-Frequency representations can be exploited for classification of unknown instantaneous frequency laws disrupted by a multiplicative noise. So, an improvement of the local SNR can be hope in a decision context.

### 3. DETECTION WITH TVHOS4

In first, we consider the basic problem of detecting the presence or absence of a signal  $\{x(n); n=0,1,\dots,N-1\}$  following the relation (4) in a set of measurement  $\{r(n); n=0,1,\dots,N-1\}$  corrupted by independent, additive, white gaussian noise with zero mean :

$$\begin{cases} H_0 : r(n) = b_a(n) \\ H_1 : r(n) = x(n) + b_a(n) \end{cases}$$

The detection method must enable to decide between  $H_0$  and  $H_1$  by analyzing the received signal  $r(n)$ . By analogy with the classical correlator detector, we can construct a time-frequency correlator and also a multitime-frequency correlator based on *TVHOS*. For the fourth order, the detection statistic is obtained by the inner-product of the *TVHOS4<sub>r</sub>* of the received signal  $r(n)$  and a reference *TVHOS4<sub>ref</sub>* obtained by averaging on independent realizations of  $x(n)$  or by taking the *LW4* of the polynomial phase signal in equation (4) as depicted on the figure 5.

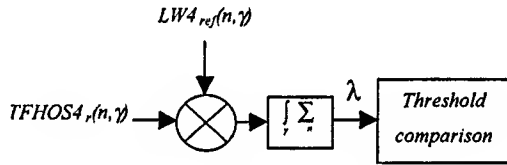


Figure 5 : TVHOS4 Based detector

#### Time equivalence and statistical behavior

Theoretical deflexion is derived in the case of *MWD4* detector. Using Moyal relation, *MWD<sub>r</sub>* detector can be written as :

$$\begin{aligned} \lambda &= \int \sum_n MWD_{4ref}(n, \gamma) \cdot MWD_{4r}(n, \gamma) d\gamma \\ &= \frac{1}{8} E \left\{ \left| \sum_n ref^2(n) \cdot r^{*2}(n) \right|^2 \right\} \end{aligned}$$

For both  $b_m(n)$  and  $b_a(n)$  white, zero mean and gaussian, we can express the deflection at the detector output by :

$$D = \frac{2[E\{\lambda|H_1\} - E\{\lambda|H_0\}]^2}{V\{\lambda|H_0\} + V\{\lambda|H_1\}}$$

where  $E(\lambda/H_i)$  and  $V(\lambda/H_i)$  are respectively mathematical expectation and variance of the detector's output under  $H_i$  hypothesis. So, we obtain the following results for the *MWD<sub>r</sub>* detector ( $D_1$ ) and for the *WVD* ( $D_2$ ) (which is only an energy detector) :

$$\begin{aligned} D_2 &= \frac{2N(\sigma_{ba}^4 + 4\sigma_{ba}^2\sigma_{bm}^2 + 4\sigma_{bm}^4)}{3\sigma_{ba}^4 + 2\sigma_{ba}^2\sigma_{bm}^2 + 2\sigma_{bm}^4} \\ D_1 &= \frac{2(\sigma_{bm}^8(N^3 + 4N^2 + 4N) + 16N\sigma_{bm}^4\sigma_{ba}^4 + 8N(N+2)\sigma_{bm}^6\sigma_{ba}^2)}{\sigma_{bm}^8(8N^2 + 44N + 90) + \sigma_{bm}^2\sigma_{ba}^6(8N + 48) + \sigma_{ba}^2\sigma_{bm}^6(8N^2 + 80N + 192) + \sigma_{ba}^4\sigma_{bm}^4(4N^2 + 34N + 124) + \sigma_{ba}^8(8N + 64)} \end{aligned} \quad (6)$$

If we draw the 2 expressions of the equation (6) versus multiplicative and additive noise standard deviation, (figure 6 and 7) we can conclude that *WVD* performs a better detection than *MWD<sub>r</sub>* for the signal (4). We illustrated this theoretical results through the detection of a signal following the relation (4) with  $\Phi'(n) = 8.92 \cdot 10^{-3}n + 0.314$ .  $b_m(n)$  is a real gaussian noise process with a bandwidth equal to 0.1 in normalized frequency and  $b_a(n)$  is a white complex circular gaussian noise. The SNR is defined by  $\sigma_{bm}/\sigma_{ba}$  and was taken to -10 dB. The results are averaged on 20 independent realizations. The ROC curves clearly indicate ( Figure 8 ) the better results obtained by the *WVD* detector and they also confirm the statistical results mentioned above.

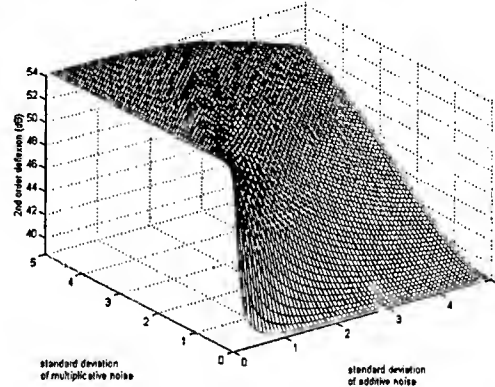


Figure 6 : Theoretical deflexion *WVD*

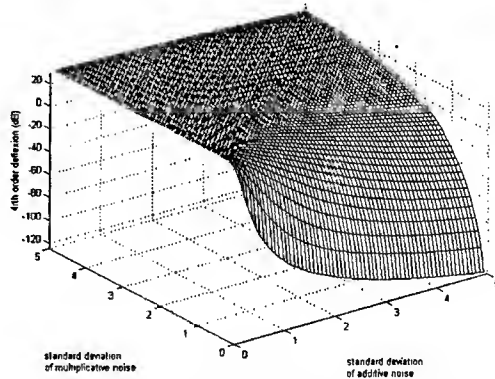


Figure 7 : Theoretical deflexion for *MWD<sub>r</sub>*

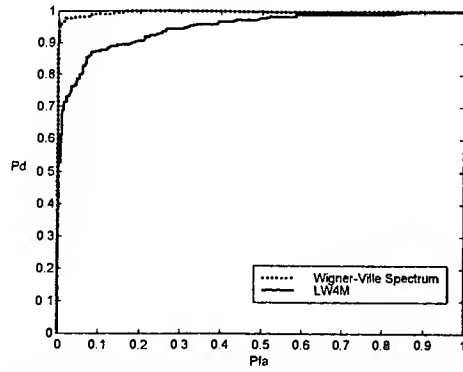


Figure 8 : ROC curves for *MWD4* and *WVD* detector

#### 4. CLASSIFICATION WITH TVHOS4

If we extend the decision rule to generalized detection with multiple hypotheses tests, we obtain in the *TVHOS4* space :

$$\begin{cases} H_1 : TVHOS4_r(n, \gamma) = TVHOS4_{x_1(t)+b_a}(n, \gamma) \\ \vdots \\ H_L : TVHOS4_r(n, \gamma) = TVHOS4_{x_L(t)+b_a}(n, \gamma) \end{cases}$$

where *TVHOS4* can be *MWD4* or *CWD4* representations. So, the classification scheme is a bank of *L* “*TVHOS4*–energy” compensated detectors (to ensure normalization) as presented on the figure 9.

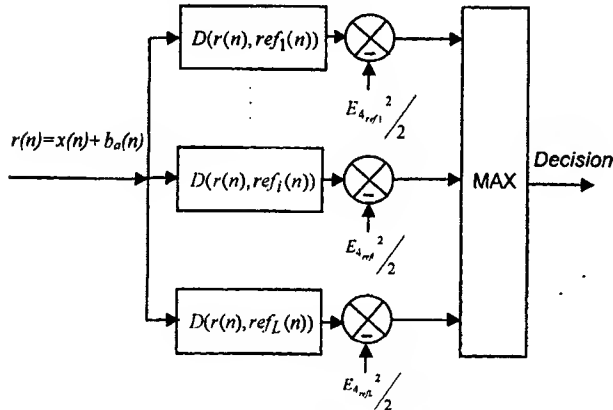


Figure 9 : *TVHOS4* based classifier

The output of each detector can be view as a special case of “minimum distance”:

$$d_i^2 = \int_{n, \gamma} (TVHOS4_r(n, \gamma) - TVHOS4_{ref_i}(n, \gamma))^2 dnd\gamma$$

after some calculus, we obtain :

$$d_i^2 = -2 \int_{n, \gamma} TVHOS4_r(n, \gamma) TVHOS4_{ref_i}(n, \gamma) dtd\gamma + E_{4r}^2 + E_{4ref_i}^2$$

and the decision rule is :

$$\eta = \arg \max_{ref_i} [D(r(n) / ref_i(n)) - \frac{1}{2} E_4(ref_i(n))]$$

where  $D(r(n), ref_i(n)) = \langle TVHOS4_r, TVHOS4_{ref_i} \rangle$ .

#### 5. CLASSIFICATION OF NEIGHBORING FREQUENCY MODULATIONS.

Performances were illustrated for classification of two neighboring instantaneous frequency modulations laws (IFL) in a context of multiplicative and additive noise. The aim of this simulation is to separate two signals following the relation (5) with  $a=10/\pi$ ,  $b = 0.125$  or  $0.130$ . For different *SNR* values and for three multiplicative noise bandwidths (indicated on the figures in normalized frequency), 1000 Monte-Carlo runs were performed. All the representations were estimated by averaging on ten independent realizations. In each case, the percentage of non-correct signal classification (error probability) is plotted on the next figures.

##### Simulations results:

First, we can show (figures 10,11 and 12) that when the bandwidth of  $b_m(n)$  increase, the two fourth order classifier gives much better results down to  $-3$  and  $-6$  dB. However we can note that the performances of two *fourth order classifier* are better than *WVD-classifier* at high *SNR*. This remark is valid for any multiplicative noise frequency bandwidth, but the results indicate that this tendency is emphasized for large multiplicative noise bandwidth. Figure 13 represents the behavior of the *WS based classifier* and the *CWD4 based classifier* versus multiplicative noise bandwidth and *SNR*. We can see that the *TVHOS4-Classifiers* are less susceptible than *WVS Classifier* to the multiplicative noise bandwidth. *CWD4 Classifier* yields very good performances and a quasi-total immunity to multiplicative noise bandwidth. Other simulation on linear frequency modulations can be found in [8]. These results lead to the same conclusion that those presented in this part.

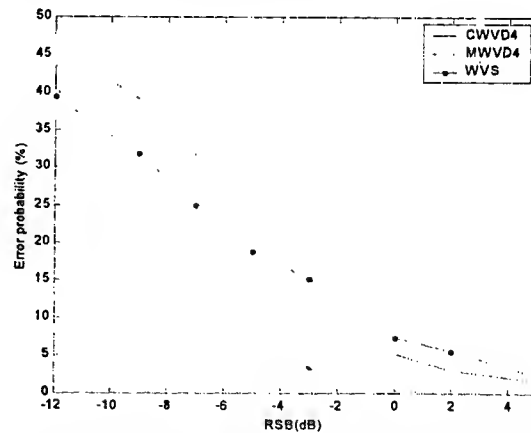


Figure 10 : Multiplicative noise bandwidth 0.02Hz

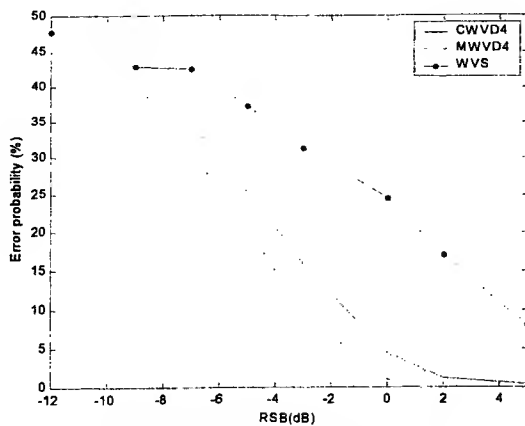


Figure 11 : Multiplicative noise bandwidth 0.07Hz

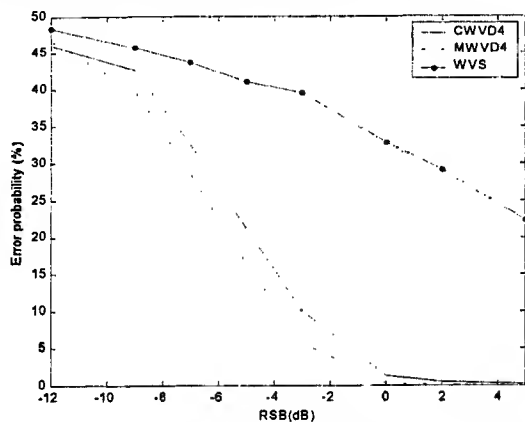


Figure 12 : Multiplicative noise bandwidth 0.12Hz

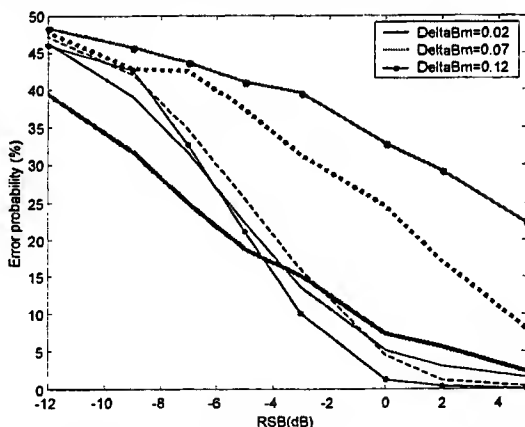


Figure 13 : WS (broad lines) , CWVD4 (fine lines)

## 6. CONCLUSION

In this paper, two new classifiers based on *TVHOS* were presented. The performances were evaluated through simulations, and we clearly show that, this approach really improves the performances to classify neighboring IFL modulation laws. Moreover, *TVHOS4* classifiers superiority is

really significant when the multiplicative noise bandwidth increases and the signal to noise ratio is sufficiently high. Performance comparison with the optimal (ML) detector is currently under consideration but is non obvious because of the non gaussianity of the signal (4). Simulation methods (MCMC) can provide an alternative solution in order to approach the theoretical optimal solution and will be presented in future works.

## 7. REFERENCES

- [1]: J. R. Fonollosa, C. L. Nikias, "Wigner Higher Order Moment Spectra : Definition, Properties, Computation and Application to Transient Signal Analysis", *IEEE Trans. On SP*, vol 40, N°1, pp 245-265, Jan 1993.
- [2]: L. Stankovic, S. Stankovic, Z. Uskokovic, "A Multi-Time Definition of the Wigner Higher Order Distribution ; L-Wigner-Distribution", in *IEEE Signal Processing Letters*, Vol. 1, n°7, July 1994.
- [3]: J. R. Fonollosa, C. L. Nikias, "Analysis of Finite Energy Signals Using Higher Order Moments-and Spectra-Based Time-Frequency Distributions", *Signal Processing*, vol 36, pp 315-328, 1994.
- [4]: L. Stankovic, S. Stankovic, Z. Uskokovic, "Time frequency signal analysis", *research monograph*, University of Montenegro, Epsilon and Montenegropulic, 1994.
- [5]: B. Boashash, B. Ristic, " Polynomial time-frequency distributions and time-varying higher order spectra : Application to the analysis of multicomponent FM signals and the treatment of multiplicative noise", *Signal Processing*, Vol 67, N°1, pp 1-23, 1998.
- [6]: H. L. Van Trees, " Detection, Estimation and Modulation Theory, Radar, Sonar, Signal Processing and Gaussian Signal in Noise.", Krieger Publishing Company, Malabar Florida, 1992.
- [7]: P. Maragos, J. Kaiser and T. Quatieri, "Energy Separation in Signal Modulation whith Application to Speech Analysis.", *IEEE trans on Signal Processing*, vol 41, pp 3024-3051, 1993.
- [8]: M. Colas, G. Gelle, G. Delaunay, " Sliced Higher order Time Frequency Representations for detection and Classification. Acoustical Diagnosis Application to faulty fan belts", in *Proc. of IEEE SPW-HOS'99*, Cesarea, Israel, pp 183-186, 1999.
- [9]: G. Gelle, M. Colas, G. Delaunay, " Multitime Frequency Classifiers for Neighboring Modulations in a Multiplicative Noise environment", in *Proc. of ICASSP'00*, Istambul, Turkey, June 2000.

# NDA PLL DESIGN FOR CARRIER PHASE RECOVERY OF QPSK/TDMA BURSTS WITHOUT PREAMBLE

*Junghoon Lee*

COMSAT Laboratories  
22300 COMSAT Drive  
Clarksburg, MD 20871, USA

## ABSTRACT

A non-decision-aided (NDA) PLL method, which recovers the carrier phase of QPSK/TDMA bursts without preamble by iterative processing, is presented. The characteristics of the phase detector in the loop are examined and the results show that the NDA PLL exhibits similar performance to a 4<sup>th</sup> power PLL. The phase error performance was simulated and the results indicate that the proposed NDA PLL is applicable to recovering carrier phase of QPSK/TDMA bursts of an order of 100 symbols or more in length.

## 1. INTRODUCTION

In satellite communications, the QPSK modulation technique has been widely used in conjunction with time division multiple access (TDMA) mode of operation. Carrier phase recovery of bursts is a major design issue in TDMA systems. While a phase-locked loop (PLL) circuit can generally provide good carrier phase tracking performance, it typically takes a long time to acquire the carrier phase. Therefore, a long preamble is necessary for carrier phase recovery. This paper presents a non-decision-aided (NDA) PLL method, which recovers the carrier phase of a QPSK/TDMA burst without a preamble by iterative processing, thereby reducing the burst overhead.

Section 2 describes the NDA PLL. A set of equations to update the loop are presented. Phase detector characteristics of the NDA PLL are examined in section 3. Section 4 presents and discusses performance simulation results for the NDA PLL.

## 2. NDA PLL DESIGN

The block diagram of the receiver is shown in Figure 1. Here, the local oscillator does not track the phase of the received carrier. Instead, the carrier phase is recovered by subsequent processing which uses in-phase and quadrature sampled values with reference to the local oscillator phase. The carrier phase recovery process includes calculating the phase of each sampled value ( $X_n$ ,  $Y_n$ ), digital PLL processing for estimating the carrier phase, and rotating the phase of sampled values.

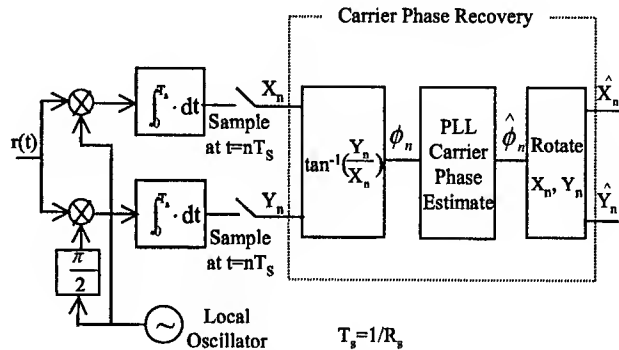


Figure 1. Receiver Block Diagram

A block diagram of the proposed NDA PLL is depicted in Figure 2. It is derived by adding a saw-tooth nonlinear function to the second-order PLL presented in [1]. The overall design concept is similar to that of the timing recovery circuits in [2], in a sense that a saw-tooth nonlinear function is used to eliminate the QPSK modulation and a feedback structure is used to unwrap and filter the phase estimate. It is noted that a filtering method in [2] resembles a first-order PLL, which can be obtained by putting  $K_2=0$  in Figure 2. Parameters of the NDA PLL are given by

$$K_1 = 2\zeta \left( \frac{2\pi f_n}{R_s} \right) \text{ and } K_2 = \frac{1}{K_1} \left( \frac{2\pi f_n}{R_s} \right)^2; \text{ and } f_n \text{ and } \zeta \text{ respectively}$$

denote the loop natural frequency and damping factor of the PLL; and  $R_s$  denotes symbol rate. The loop is updated using the following equations (1)-(4):

$$\phi_{e,n} = \phi_{a,n} - \frac{\pi}{2} \text{Int} \left( \frac{2\phi_{a,n}}{\pi} + 0.5 \right) \quad (1)$$

$$S_n = \dot{\phi}_n + K_1 \cdot \phi_{e,n} \quad (2)$$

$$\dot{\phi}_{n+1} = \dot{\phi}_n + K_1 \cdot K_2 \cdot \phi_{e,n} \quad (3)$$

$$\hat{\phi}_{n+1} = \hat{\phi}_n + S_n \quad (4)$$

where  $\text{Int}(\cdot)$  rounds a number down to the nearest integer. From Figure 2, one can note that the PLL provides the carrier phase estimate  $\hat{\phi}_n$  as well as the frequency offset estimate  $\dot{\phi}_n$ . Therefore, using those estimates, the iterative PLL processing is done as follows. By having  $\hat{\phi}_n$  and substituting  $\dot{\phi}_n$  with  $-\dot{\phi}_n$  at the end of a burst, the PLL process continues in an opposite direction. Once the PLL acquires the received signal, then the



loop parameter can be changed to get a better tracking performance. For example, the PLL loop bandwidth is reduced to achieve smaller phase error variance. In designing the PLL in Figure 2, the impact of the nonlinear phase detection, which is discussed in the next section, should be taken into consideration in addition to typical PLL design principles.

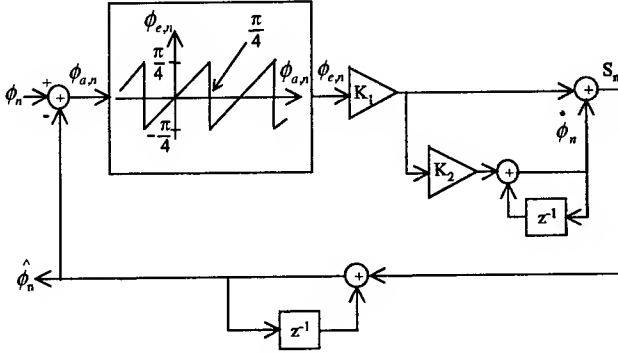


Figure 2. Block Diagram of the NDA PLL

### 3. PHASE DETECTOR CHARACTERISTICS OF THE NDA PLL

The phase detection process of the NDA PLL includes  $\tan^{-1}(\cdot)$  and saw-tooth nonlinear functions. Let us define the phase detector non-linearity as  $g[\varphi(t) + \theta_n(t)]$ , where  $\varphi(t) = \phi_n(t) - \hat{\phi}_n(t)$  and  $\theta_n(t)$  denotes input phase noise. Assuming that the phase of the input carrier varies much more slowly than the input phase noise and the bandwidth of the PLL is much smaller than the symbol rate  $R_s$ , then the phase error  $\varphi(t)$  varies much more slowly than the input phase noise  $\theta_n(t)$ . Therefore, the low frequency component of the detector output will represent the phase detector characteristics; thus the characteristics can be evaluated from the expectation

$$g'(\varphi) = E \{ g[\varphi + \theta_n] | \varphi \} \quad (5)$$

$g'(\varphi)$  of the NDA PLL is obtained by simulation and shown in Figure 3. As shown in the analysis on phase detectors in [4], the effective phase detector output is suppressed as  $(E_b/N_0)$  becomes low. The shape of the phase detector characteristics is similar to a 4<sup>th</sup> power or four-phase remodulator PLL, which is illustrated in Chapter 11 of [5].

Let's consider the signal to noise ratio (SNR) at the phase detector output. For a linear phase detector, the SNR at the phase detector output can be given by

$$SNR_{PD,L} = \frac{1}{\sigma_{PD,L}^2} = 4 \cdot \left( \frac{E_b}{N_0} \right) \quad (6)$$

where  $\sigma_{PD,L}^2$  denotes phase variance at the phase detector output.

If we define SNR of the NDA PLL as  $SNR_{PD,NDA}$ , the ratio of the SNR with reference to  $SNR_{PD,L}$  is given by

$$R_{O,NDA} = \frac{SNR_{PD,NDA}}{SNR_{PD,L}} \quad (7)$$

$R_{O,NDA}$  is obtained by simulation and shown in Figure 4, along with the SNR ratio of a 4<sup>th</sup> power PLL,  $R_{O,4th}$ , which is given in [6].

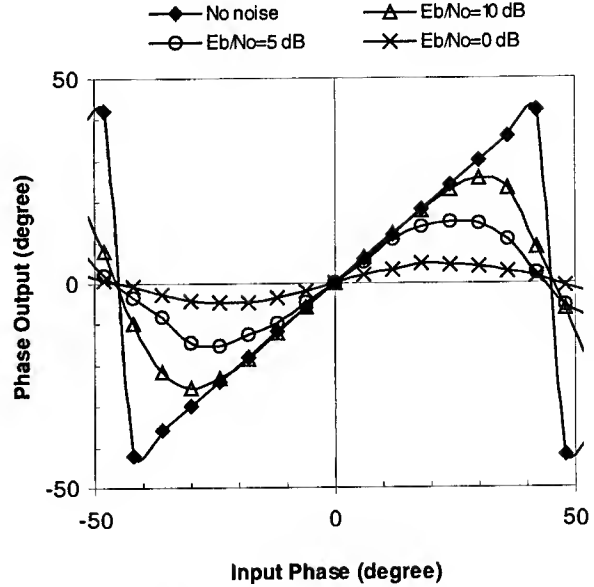


Figure 3. Phase Detector Characteristics of the NDA PLL

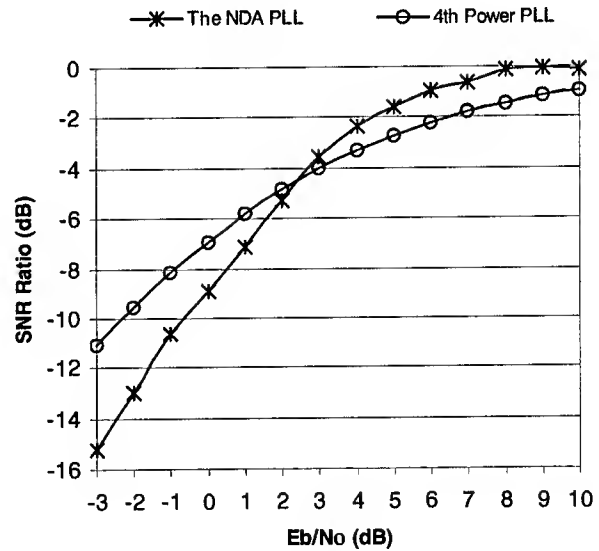


Figure 4. SNR Ratios of Phase Detectors

#### 4. SIMULATION RESULTS AND DISCUSSION

Based on the analysis in [5], phase variance  $\sigma_{O,L}^2$  of a second order PLL with linear phase detector is given by

$$\frac{1}{\sigma_{O,L}^2} = 2 \left( \frac{E_b}{N_o} \right) \left( \frac{R_s}{f_n} \right) \left\{ \frac{4\zeta}{\pi(1+4\zeta^2)} \right\} \quad (8)$$

Phase variance of a 4<sup>th</sup> power PLL and the NDA PLL can be estimated using the SNR ratios obtained in the previous section. Including the calculated tracking error performances for the linear PLL, 4<sup>th</sup> power PLL, and NDA PLL by  $\sigma_{O,L}$ ,  $\sigma_{O,L}/\sqrt{R_{O,4th}}$ , and  $\sigma_{O,L}/\sqrt{R_{O,NDA}}$ , respectively, Figures 5 and 6 illustrate the simulation results on the steady-state phase tracking error performance of the NDA PLL. At high SNR the simulated NDA PLL performance closely matches the calculated performance based on the SNR ratio. At low SNR the NDA performance becomes worse than the calculated based on the SNR ratio. It is due to the loop threshold<sup>[7], Chapter 6</sup>, that is, occurrence of cycle slips. The phase detector characteristics of the NDA PLL are similar to the 4<sup>th</sup> power PLL. The Nth power PLL elevates loss of lock by approximately  $20 \log_{10} N$  dB<sup>[4], Chapter 11</sup>. Assuming that the loop threshold point of a second order PLL is 8 dB as illustrated in Chapter 6 of [7], loop threshold of the NDA PLL becomes 20 dB, or a phase error standard deviation of 5.7 degrees. From Figures 5 and 6, comparing the simulated results of the NDA PLL with the calculated ones, it is noted that the threshold approximately agrees to the 5.7 degrees point.

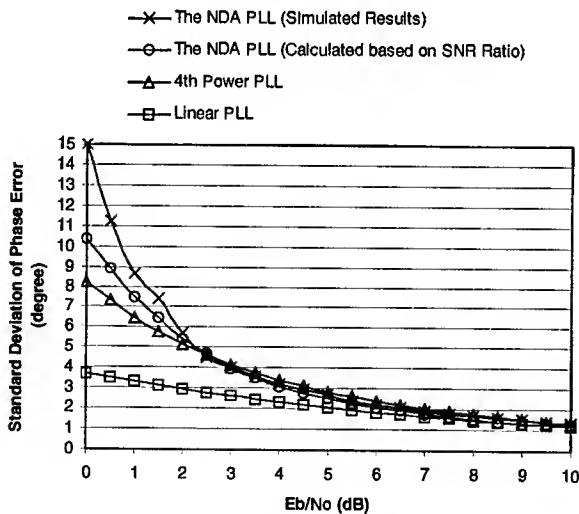


Figure 5. Steady-state Tracking Performance ( $R_s/f_n=400$ ,  $\zeta=0.707$ , no frequency offset)

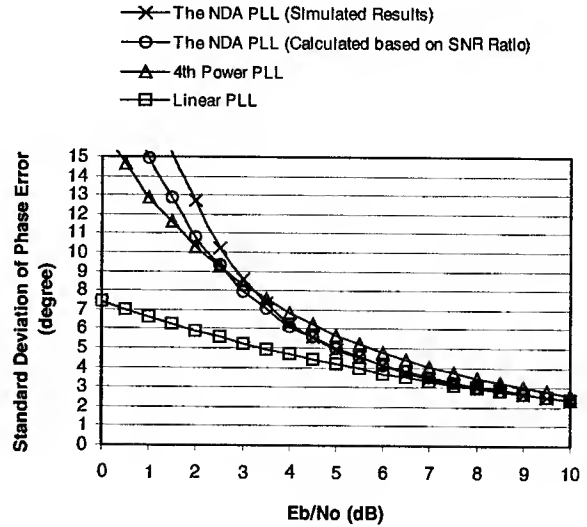


Figure 6. Steady-state Tracking Performance ( $R_s/f_n=100$ ,  $\zeta=0.707$ , no frequency offset)

Now, consider initial acquisition performance. During initial acquisition, a second-order PLL will exhibit a phase transient depending on the initial phase error and frequency offset. When the phase transient reaches dynamic range of the NDA PLL, that is approximately  $1/4$  of a linear PLL or  $\pi/8$  at low SNR, the acquisition process is disturbed and the loop might go to a random status. So the initial acquisition transient might occur again. Therefore, in the NDA PLL design, care should be taken such that the initial phase error and frequency offset should be kept sufficiently small for the loop to acquire a signal within a proper time, for example, a burst period. The following simulation results illustrate that it is feasible to design the NDA PLL in such a way that the loop properly acquires carrier phase of bursts of 160 symbols or more in length.

The phase error performance of iterative burst processing using the NDA PLL is simulated and the results are illustrated in Figures 7 and 8. The block phase estimation method in [3] is used to get the initial value. In Figure 7, the PLL parameters of  $R_s/f_n=200$  and  $\zeta=0.707$  are used for the initial PLL processing in the forward direction.  $R_s/f_n=400$  and  $\zeta=0.707$  are used for the subsequent PLL processing in the backward direction. In Figure 8, the PLL parameters of  $R_s/f_n=100$  and  $\zeta=0.707$  are used both for the initial PLL processing in the forward direction and for the subsequent PLL processing in the backward direction. The simulation results show that the same performance as steady-state tracking phase error can be obtained when the carrier frequency offset is small. It implies that for the given parameters, the loop acquires the phase at the initial PLL processing in the forward direction.

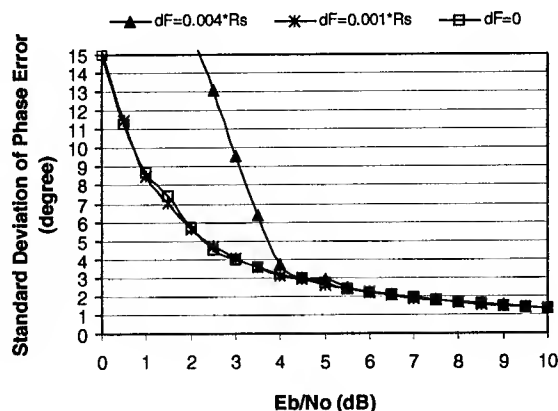


Figure 7. Simulated Phase Error Performance of Iterative Burst Processing Using the NDA PLL ( $dF$ : carrier frequency offset, burst length=640 symbols)

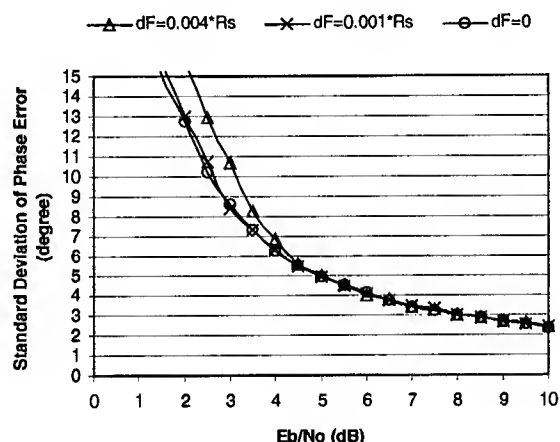


Figure 8. Simulated Phase Error Performance of Iterative Burst Processing Using the NDA PLL ( $dF$ : carrier frequency offset, burst length=160 symbols)

As the burst length becomes shorter, it becomes harder to adjust the PLL parameter  $R_s/f_n$  such that the loop can properly acquire the phase within a burst period. In that respect, further simulation results indicate, although not shown here, that the NDA PLL is applicable to a burst length of an order of 100 or more.

## 5. CONCLUSION

An NDA PLL method, which recovers the carrier phase of QPSK/TDMA bursts without preamble by iterative processing, is presented. The characteristics of the phase detector in the loop are examined and the results show that the NDA PLL exhibits similar performance to a 4<sup>th</sup> power PLL. The phase error performance was simulated and the results indicate that the proposed NDA PLL is applicable to recovering carrier phase of QPSK/TDMA bursts of an order of 100 symbols or more in length.

## 6. REFERENCES

- [1] S. A. Rhodes and S.I. Sayegh, "Digital Onboard Processing for Reception of an Uplink Group of TDMA/QPSK Channels," *Proceedings of 8<sup>th</sup> International Conference on Digital Satellite Communications*, pp. 845-852, Guadeloupe, F. W. I., April 1989.
- [2] M. Oerder and H. Meyr, "Digital Filter and Square Timing Recovery," *IEEE Transactions on Communications*, vol. 36, No.5, pp. 605-612, May 1988.
- [3] A. J. Viterbi and A. M. Viterbi, "Nonlinear Estimation of PSK-Modulated Carrier Phase with Application to Burst Digital Transmission," *IEEE Transactions on Information Theory*, vol. IT-29, pp. 543-551, July 1983.
- [4] W. Rosenkranz, "Phase-Locked Loops with Limiter Phase Detectors in the Presence of Noise," *IEEE Transactions on Communications*, vol. 30, No.10, pp. 2297-2304, October 1982.
- [5] F. M. Gardner, *Phaselock Techniques*, John Wiley & Sons, New York, 1979.
- [6] S. A. Butman and J. R. Lesh, "The Effects of Bandpass Limiters on  $n$ -Phase Tracking Systems," *IEEE Transactions on Communications*, vol.25, No.6, pp. 569-576, June 1977.
- [7] H. Meyr and G. Ascheid, *Synchronization in Digital Communications*, John Wiley & Sons, 1990.

# AN OPTIMIZED MULTI-TONE CALIBRATION SIGNAL FOR QUADRATURE RECEIVER COMMUNICATION SYSTEMS

Roger A. Green

Elec. & Comp. Eng. Dept., North Dakota State University,  
P.O. Box 5285, Fargo, ND 58105-5285  
Phone: (701) 231-1024, e-mail: [green@prairie.nodak.edu](mailto:green@prairie.nodak.edu)

## ABSTRACT

Communication systems are subject to stringent image rejection requirements. Thus, accurate and regular field calibration is important. Regression techniques are effective in the calibration of quadrature receiver systems. These techniques require the transmission or injection of a calibration signal to estimate potentially frequency-dependent errors. Existing regression-based methods use nonlinear models with signals that calibrate only one frequency at a time. This paper recasts the problem in terms of linear regression and develops an optimized multi-tone calibration signal for quadrature receiver communication systems. Linear regression ensures closed-form solutions that can be computed in real-time by using adaptive filtering techniques. Simulations demonstrate the advantages of the multi-tone signal: simultaneous multi-frequency calibration and minimal interference with information bearing communication channels. At the same time, the benefits of regression-based calibration are also realized: modest model assumptions, effective performance assessment, and accommodation of non-uniformly sampled or missing calibration data.

Keywords: System identification and calibration, Signal processing for communications

## 1. INTRODUCTION

Green *et al.* developed a nonlinear regression (NLR) -based method to calibrate gain and phase mismatch between the in-phase (I) and quadrature (Q) branches of a quadrature receiver [4, 5]. The method is effective and allows reliable error assessment. Due to the nonlinear models, however, real-time implementation can be difficult. With trigonometric manipulation the technique can be recast in terms of linear regression (LR). The technique requires either the transmission or injection of a calibration signal to estimate potentially frequency-dependent errors. The use of a transmitted calibration signal is particularly advantageous since this permits regular field calibration of the receiver without adding complex hardware.

Figure 1 illustrates a standard quadrature receiver. For simplicity, the antenna is assumed to be omni-directional. The gain and phase errors are modeled by the impulse response functions  $h_I$  and  $h_Q$ , for the I and Q branches respectively. This approach allows errors to be frequency dependent as shown by the frequency-domain representations of

the response functions:  $H_I(\omega) = G_I(\omega) \exp(j\psi_I(\omega))$  and  $H_Q(\omega) = G_Q(\omega) \exp(j\psi_Q(\omega))$ . Although the errors are modeled as frequency dependent, it is reasonable to assume that the gain and phase errors are approximately constant over narrow frequency bands.

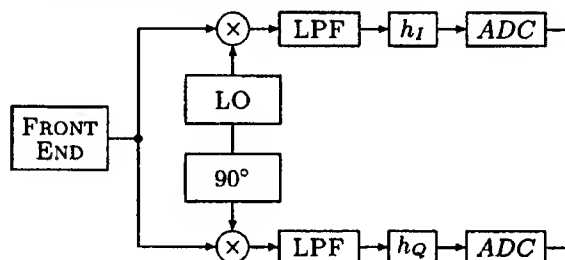


Figure 1: I/Q Receiver with Gain and Phase Imbalances

Although other calibration procedures exist (see [1], [10], [8], and [7]), regression-based calibration techniques provide several advantages over these methods including modest model assumptions, allowance of non-uniformly sampled or missing calibration data, effective performance assessment, and model flexibility. Furthermore, regression-based techniques can calibrate receivers in normal operating environments. That is, test signals are transmitted to receivers according to field design; while direct test signal injection is supported, it is not necessary.

In the context of communication systems, several LR properties are particularly important. First, real-time calibration must be accommodated. Closed-form solutions are straight-forward with LR, and sequential parameter updates are possible using adaptive filters. Second, the capability to calibrate using a transmitted signal is important to avoid the restrictive addition of hardware. Third, the ability to accurately monitor estimator performance is beneficial. Modern mobile communications, for example, desire around 70 dB of image rejection; for quadrature receivers, this corresponds to roughly  $1/20^\circ$  of allowable phase deviation between the I and Q branches. System performance is monitored using parameter inferences and confidence intervals.

In [4] and [5], two calibration signals are considered: a pure sinusoid and a double sideband suppressed carrier signal. Unfortunately, both signals suffer from the significant disadvantage that only one frequency is calibrated at

a time. This impedes the task of calibrating a receiver over the frequency range of operation. Practical implementation requires a signal that admits simultaneous calibration over multiple frequencies. Through proper selection of frequencies and optimization of relative phase, a multi-tone signal can be constructed that is suitable for practical regression-based calibration.

## 2. LR-BASED CALIBRATION

Observations that follow a LR model are expressed according to  $Y = X\beta + \epsilon$ , where  $Y$  is the  $N$ -by-1 vector of observed values;  $X$  is the known  $N$ -by- $P$  predictor matrix;  $\beta$  is the  $P$ -by-1 unknown parameter vector; and  $\epsilon$  is the  $N$ -by-1 vector of additive, zero-mean, uncorrelated noise. The LR model does not require a particular sampling scheme. Thus, non-uniformly sampled or missing data do not affect the method.

Given a sufficient number of observations, unique closed-form solutions for  $\beta$  generally exist, although problems such as multicollinearity occasionally arise. Using the solution for deterministic least squares, the unknown parameter vector is estimated according to

$$\hat{\beta} = (X'X)^{-1} X'Y, \quad (1)$$

where  $X'$  is the matrix transpose of  $X$  and  $(-1)$  designates the matrix inverse operation.

Equation (1) is well suited for post-processing or off-line calibration. In cases where real-time calibration is desired, adaptive filters can be utilized. Preliminary results using the Least Mean Square (LMS) algorithm and the Recursive Least Squares (RLS) algorithm are promising [6]. Currently, gain and phase errors are modeled as constants with respect to time. Although additional research is needed to address non-stationarities such as component drift, methods such as the Exponential Forgetting Window (EFW)-RLS algorithm are designed just for such conditions.

There exist several effective methods to ascertain LR estimator performance. Provided independent, normal errors and a reasonable sample size,  $\alpha$ -level confidence intervals are

$$[\hat{\beta}_p - t_{N-P} \{\hat{\beta}_p\}, \hat{\beta}_p + t_{N-P} \{\hat{\beta}_p\}], \quad (2)$$

where  $t_{N-P}(1 - \alpha/2)$  is the  $(1 - \alpha/2)$  100 percentile of the student- $t$  distribution with  $(N - P)$  degrees of freedom. The sample variance-covariance matrix of  $\beta$  is given by  $s^2(\beta) = (Y' - \hat{\beta}'X')Y(X'X)^{-1}/(N - P)$ .

In communications systems, assumptions about  $\epsilon$  are likely violated. For example, the error term includes signal content from the communication channels themselves; and these signals rarely exhibit time independence and normality. Most violations, however, have minimal impact on estimation since data sets are generally large by conventional standards. Data correlations will not bias estimates, but may erroneously shrink confidence intervals. In these cases, conservative specifications should be used. Details are in [4].

Resampling techniques, such as the jackknife or the bootstrap, have no underlying assumptions regarding sample size or normality of errors, and they provide effective

alternatives to (2). This is particularly true in cases where the desired parameters are functions of  $\beta$  and not the individual terms  $\beta_p$  themselves. Details for jackknife and bootstrap inferences are given in [4].

## 3. OPTIMIZED MULTI-TONE SIGNAL

A multi-tone calibration signal  $m(t)$  can be constructed using a superposition of real sinusoids

$$m(t) = \sum_{k=1}^K M_k \cos(\omega_k t + \theta_k), \quad (3)$$

where  $M_k$  and  $\theta_k$  establish the relative magnitude and phase of each sinusoid component [3]. By restricting our basis set to sinusoids, frequency content is easily controlled. To simplify calibration, it is most sensible to constrain all gains to be equal,  $M_k = M$  for all  $k$ . The final value  $M$  is chosen to reflect the desired signal power. Since the power of  $m(t)$  is distributed over frequency, individual components possess relatively low power, which helps minimize channel interference by the calibration signal. Construction of the calibration signal, then, requires determination of  $K$ ,  $\theta_k$ , and  $\omega_k$  for all  $k$ .

The number of calibration frequencies  $K$  should be sufficient to cover the frequency range of interest. This parameter will vary from application to application. In some cases,  $K$  may be quite large. This, in turn, can cause serious problems in dynamic range [3]. From (3) it is clear that the worst case occurs when all phases are zero,  $\theta_k = 0$  for all  $k$ . The resulting signal has nearly all of its power concentrated in very narrow time slots. Even distribution of signal power over time improves dynamic range. To this end, the phase terms  $\theta_k$  are chosen to minimize the amplitude of the calibration signal,

$$\min_{\theta_k} \left( \max_t (|m(t)|) \right). \quad (4)$$

The phase optimization in (4) is a nonlinear problem. To compute  $\theta_k$ , define  $m_{\max} = \max_t (|m(t)|)$  and let  $t_{\max}$  be the time when this maximum occurs. That is,  $m_{\max} = |m(t_{\max})|$ .

From (3), the first and second partial derivatives of  $m(t_{\max})$  with respect to  $\theta_k$  are given by

$$\frac{\partial m(t_{\max})}{\partial \theta_k} = -M_k \sin(\omega_k t_{\max} + \theta_k) \quad (5)$$

and

$$\frac{\partial^2 m(t_{\max})}{\partial \theta_k^2} = -M_k \cos(\omega_k t_{\max} + \theta_k) \quad (6)$$

When set equal to zero, (5) identifies critical points. Equation (6) determines whether these points are maxima, minima, or points of inflection.

To simplify computation of the first two partial derivatives of  $m_{\max}$ , (5) and (6) are combined using Euler's identity

$$\frac{\partial^2 m_{\max}}{\partial \theta_k^2} + j \frac{\partial m_{\max}}{\partial \theta_k} = -\text{sgn}(m(t_{\max})) M_k e^{j(\omega_k t_{\max} + \theta_k)}. \quad (7)$$

Suitable  $\theta_k$  are computed using gradient-descent type algorithms, which are made efficient by using (7).

It is critical that the frequencies  $\omega_k$  are chosen so that the resulting calibration signal is periodic. That is,  $\omega_k$  must allow  $m(t) = m(t + T)$  for all  $t$  and at least one finite value of  $T$ . Without periodicity, phase optimization of the calibration signal is meaningless, and the resulting calibration signal will necessarily display "bad" behavior. One simple way to ensure that the calibration signal is periodic is to choose  $n_i\omega_i = n_j\omega_j$  for  $i = 1, 2, \dots, K$  and  $j = 1, 2, \dots, K$  and where  $n_i$  and  $n_j$  are integers. It is also convenient to restrict  $\omega_k > 0$  for all  $k$ . In some cases, I/Q mismatch varies slowly with frequency. In these cases, it is sensible to place  $\omega_k$  between information-bearing frequency bands. This will help minimize interference between the calibration signal and the communication channels.

Quadrature receivers operate by mixing the received signal with a local oscillator of frequency  $\omega_{LO}$ , as shown in Figure 1. This operation frequency shifts the received signal. Thus, it is necessary to translate the frequencies  $\omega_k$  of the optimized signal  $m(t)$  by  $|\omega_{LO}|$  prior to transmission. That is, the transmitted signal is obtained from 3 by replacing  $\omega_k$  with  $\tilde{\omega}_k = \omega_k + |\omega_{LO}|$  for all  $k$ . This ensures that the calibration signal, once demodulated, maintains the optimized signal with content at the desired frequencies  $\omega_k$ . Optimization of  $m(t)$  should not be done using  $\tilde{\omega}_k$  since frequency translation during demodulation otherwise disrupts the phase optimization.

#### 4. SYSTEM MODEL

The optimized multi-tone test signal  $m(t)$  is well suited for LR-based quadrature receiver calibration. Although  $m(t)$  is known for a given application, the received signal  $\tilde{m}(t)$  has unknown gain  $G$  and unknown phase-shift  $\psi$ , which correspond to channel attenuation and independent operation of the receiver's local oscillator, respectively.

Following analog-to-digital conversion, the received signal is given by

$$\tilde{m}(t) = \sum_{k=1}^K [I_I X_1(t, \omega_k) \beta_1(\omega_k) + I_I X_2(t, \omega_k) \beta_2(\omega_k) + I_Q X_1(t, \omega_k) \beta_3(\omega_k) + I_Q X_2(t, \omega_k) \beta_4(\omega_k)] + \varepsilon(t), \quad (8)$$

where  $X$  are the predictor variables,  $\beta$  are the unknown parameters, system noise is designated by  $\varepsilon(t)$ , and each  $I$  serves as an indicator function. The indicator functions identify observations from different branches and improve estimator performance [4, 5]. In this case,  $I_I = 1$  and  $I_Q = 0$  when representing data from the in-phase branch;  $I_I = 0$  and  $I_Q = 1$  when representing data from the quadrature branch.

The unknown parameters are estimated using (8) and the LR procedure described in Section 2. Recommendations in [4] regarding system sampling should be followed to ensure good parameter estimates. Essentially, the  $\omega_k t$  samples modulo  $2\pi l$ , where  $l$  is any integer, should not be tightly clustered.

In (8), the predictor variables are given by

$$\begin{aligned} X_1(t, \omega_k) &= \cos(\omega_k t + \theta_k) \\ X_2(t, \omega_k) &= \sin(\omega_k t + \theta_k), \end{aligned} \quad (9)$$

and the unknown parameters are given by

$$\begin{aligned} \beta_1(\omega_k) &= G_I(\omega_k) \cos(\alpha_I(\omega_k)) \\ \beta_2(\omega_k) &= G_I(\omega_k) \sin(\alpha_I(\omega_k)) \\ \beta_3(\omega_k) &= G_Q(\omega_k) \sin(\alpha_Q(\omega_k)) \\ \beta_4(\omega_k) &= -G_Q(\omega_k) \cos(\alpha_Q(\omega_k)). \end{aligned} \quad (10)$$

Here, the grouped gain parameters  $G_I(\omega_k)$  and  $G_Q(\omega_k)$  include contributions from the unknown test signal gain as well as the unknown frequency-dependent gains of each individual branch. Similarly, the grouped parameters  $\alpha_I(\omega_k)$  and  $\alpha_Q(\omega_k)$  represent phase contributions from the test signal as well as the receiver. Individual gain and phase terms are not important; rather, it is only the relative mismatch between branches as a function of frequency that is important.

Using the I branch as reference, the relative gain mismatch as a function of frequency,  $G_{rel}(\omega_k)$ , is given by

$$G_{rel}(\omega_k) = \sqrt{\frac{\beta_3^2(\omega_k) + \beta_4^2(\omega_k)}{\beta_1^2(\omega_k) + \beta_2^2(\omega_k)}}, \quad (11)$$

and the relative phase mismatch as a function of frequency,  $\alpha_{rel}(\omega_k)$ , is given by

$$\alpha_{rel}(\omega_k) = \arctan\left(-\frac{\beta_3(\omega_k)}{\beta_4(\omega_k)}\right) - \arctan\left(\frac{\beta_2(\omega_k)}{\beta_1(\omega_k)}\right). \quad (12)$$

As alluded to earlier, although confidence intervals for  $\beta$  are straightforward, confidence intervals for  $G_{rel}(\omega_k)$  and  $\alpha_{rel}(\omega_k)$  are complicated. Bootstrap and jackknife methods are effective, but they are computationally inefficient. Thus, practical real-time computation of confidence intervals or inferences needs further investigation.

#### 5. SIMULATIONS AND RESULTS

Computer simulations help demonstrate the effectiveness of LR-based calibration using an optimized multi-tone signal. First, an optimized calibration signal is constructed with  $\omega_k = 250(2\pi k)$  for  $k = \{1, 2, 3, \dots, 16\}$ . This signal allows calibration up to a frequency of around 4-kHz. Figure 2 plots two periods of this signal. Although the optimization procedure does not produce a unique solution, the desired temporal distribution of signal energy is achieved.

For the first calibration scenario, consider the simple case when the only errors present are a deviation from the unit-gain quadrature state. Using the in-phase branch as reference, this is equivalent to the injection of gain and phase errors  $G_{LO}$  and  $\psi_{LO}$  into the quadrature branch. The unknown parameters  $\beta$  are therefore not functions of frequency, and the number of unknowns is reduced from  $4K$  to four. Simple modification of (8) accommodates this fact.

Prior to transmission, each component in  $m(t)$  is frequency shifted by 40-kHz, to match the carrier and local oscillator frequencies. A Double Side-Band Suppressed Carrier (DSB-SC) voice signal, band-limited to 10-kHz is also

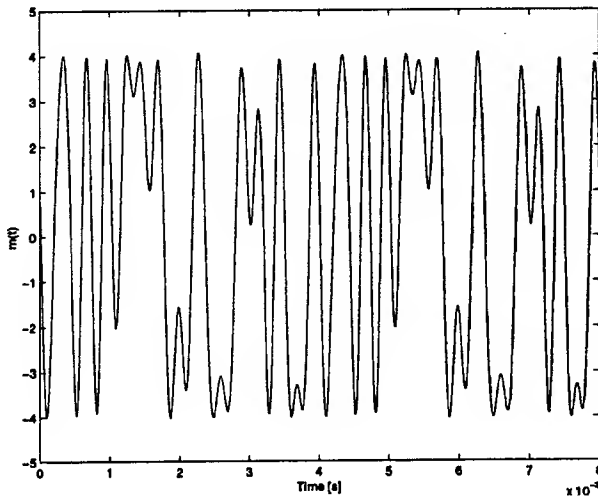


Figure 2: Optimized Calibration Signal

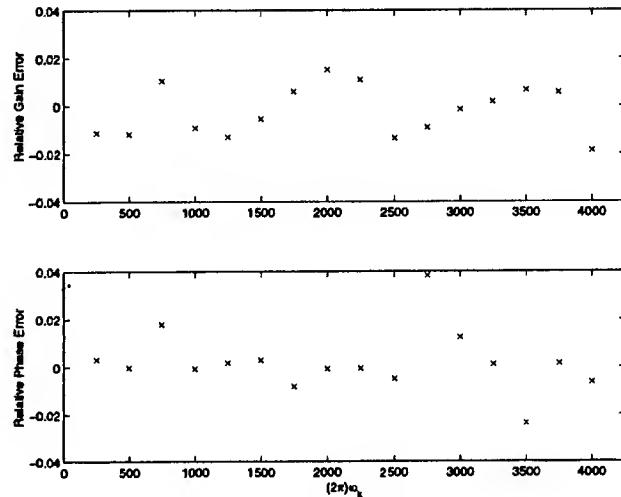


Figure 3: Relative Gain and Phase Errors

included. According to (8), the voice signal is viewed as noise  $\varepsilon(t)$ , and the Signal-to-Noise Ratio (SNR) is unity.

The received signal is split into I and Q paths, mixed, and then lowpass filtered using a 12<sup>th</sup>-order equiripple FIR filter with 10-kHz cut-off frequency and 70-dB of stopband attenuation. The Q branch is given a gain error of  $G_{LO} = 0.95$  and a phase deviation of  $\psi_{LO} = \pi/3$  radians. Receiver sampling is set to 22.05-kHz.

Using a total of 39,588 samples per branch, the parameters  $\beta$  are estimated using (1). Following transformation by (11) and (12), the gain and phase error estimates are  $\hat{G}_{LO} = 0.9498$  and  $\hat{\psi}_{LO} = 1.0296$ . This gives relative errors of -0.023 percent and -1.68 percent, respectively.

Although the correlated nature of voice violates the i.i.d. assumption typical for error terms in LR models, the calibration procedure produces good-quality estimates. This is particularly encouraging given the low SNR and relatively short data record of under two seconds.

The second calibration scenario is similar to the first. In this case, however, randomly chosen frequency-dependent errors  $\beta(\omega_k)$  are present, the LR model error  $\varepsilon(t)$  is i.i.d. Gaussian noise, and the SNR is 10 dB. Figure 3 summarizes the relative errors of the estimates. I/Q mismatch is well estimated across the frequencies  $\omega_k$  with mean relative error of around one percent.

## 6. CONCLUSIONS

An optimized multi-tone signal was developed for LR-based calibration of gain and phase mismatch in I/Q receivers. This signal permits simultaneous multi-frequency calibration and causes minimal interference with information bearing communication channels. LR estimates have closed-form solutions and permit sequential updating using adaptive filters. At the same time, the benefits of the regression-based calibration are also realized: modest model assumptions, effective parameter inferences, and accommodation of non-uniformly sampled or missing calibration data. Simulations document the effectiveness of the procedure.

## 7. REFERENCES

- [1] F. Churchill, G. Ogar, and B. Thompson, "The Correction of I and Q Errors in a Coherent Processor," *IEEE Trans. AES*, vol. AES-17, no. 1, pp. 131-137, Jan. 1981.
- [2] B. Efron, *The Jackknife, the Bootstrap, and Other Resampling Plans*, CBMS-NSF Regional Conference Series in Applied Mathematics, No. 38, Society for Industrial and Applied Mathematics, 1982.
- [3] D. C. Farden, G. Miramontes de León, and D. E. Tallman, "DSP-Based Instrumentation for Electrochemical Impedance Spectroscopy," *Electrochemical Society Proceedings Volume 99-5*, pp. 98-108, 1999.
- [4] R. A. Green, R. A. Anderson-Sprecher, and J. W. Pierre, "Quadrature Receiver Mismatch Calibration," *IEEE Transactions on Signal Processing*, Vol. 47, No. 11, Nov. 1999.
- [5] R. A. Green, "A DSB-SC Signal Model for Nonlinear Regression-Based Quadrature Receiver Calibration," *Proc. ICASSP '99*, Phoenix, Arizona, May 1999.
- [6] S. Haykin, *Adaptive Filter Theory*, 2<sup>nd</sup> Edition, Prentice Hall, Inc., 1991.
- [7] J. P. Y. Lee, "Wideband I/Q Demodulators: Measurement Technique and Matching Characteristics," *IEE Proc. Radar, Sonar Navig.*, Vol. 143, No. 5, Oct. 1996.
- [8] M. D. Macleod, "Fast Calibration of IQ Digitizer Systems," *Electrical Eng.*, pp. 41-45, 1990.
- [9] J. Neter, M. H. Kutner, C. J. Nachtsheim, and W. Wasserman, *Applied Linear Statistical Models*, 4<sup>th</sup> Edition, Richard D. Irwin, Inc., pp. 531-559, 1996.
- [10] J. Pierre and D. Fuhrmann, "Considerations in Auto-calibration of Quadrature Receivers," *Proc. ICASSP '95*, pp. 1900-1903, Detroit, Michigan, May 1995.

# A POLYNOMIAL ROOTING APPROACH FOR SYNCHRONIZATION IN MULTIPATH CHANNELS USING ANTENNA ARRAYS

Gonzalo Seco\*, A. Lee Swindlehurst†, Juan A. Fernández-Rubio\*

\*Dept. of Signal Theory & Communications, Univ. Politècnica de Catalunya, Barcelona, SPAIN.

† Dept. of Electrical & Computer Engineering, Brigham Young University, Provo, Utah, USA.

email: gonzalo/juan@gps.tsc.upc.es, swindle@ee.byu.edu

## ABSTRACT

The estimation of the delay of a known training signal received by an antenna array in a multipath channel is addressed. The effect of the co-channel interference is taken into account by including a term with unknown spatial correlation. The channel is modeled as an unstructured FIR filter. The exact maximum likelihood (ML) solution for this problem is derived, but it does not have a simple dependence on the delay. An approximate estimator that is asymptotically equivalent to the exact one is presented. Using an appropriate reparameterization, it is shown that the delay estimate is obtained by rooting a low-order polynomial, which may be of interest in applications where fast feedforward synchronization is needed.

## 1. INTRODUCTION

Time-delay estimation or timing synchronization is a key task in diverse areas, such as radar, sonar and communications. Accurate chip/symbol synchronization is especially important in systems employing time-division multiple access (TDMA) or asynchronous burst transmissions. Also, most multiuser detectors for code-division multiple access (CDMA) require reliable estimates of the users' code timings in order to operate acceptably in near-far environments [1]. In addition, Global Navigation Satellite Systems (GNSS) arouse great interest at present. In these systems, accurate time-delay estimation is fundamental, since it is the key to obtain sub-meter accuracies in location estimates.

There is a vast literature on single antenna synchronization methods for both additive white Gaussian noise (AWGN) channels and multipath channels [2]. However, the performance of these methods is limited when strong co-channel interference (CCI) is present. For this reason, an important effort is being conducted to derive time-delay estimators that make efficient use of antenna arrays in interference limited scenarios. Following an approach that has already been applied successfully to this and other problem, all the components contributing to the noise and CCI

(*i.e.*, multi-access interference (MAI), external interference, etc.) are modeled as a Gaussian term with unknown and arbitrary spatial correlation matrix [3, 4, 5]. This model allows us to develop a metric that takes the spatial characteristics of the CCI into account, and we believe that this offers an excellent trade-off between model realism and computational complexity. A more detailed description of the CCI, *e.g.* using the finite alphabet property of the MAI, may result in an improved performance, but at the expense of avoiding a simple expression for the estimator. On the other hand, several computationally attractive algorithms have been derived under the usual assumption that the CCI is spatially white [6]. However, the resulting algorithms are not suited for situations involving strong CCI. The estimation of the spatial covariance of the CCI is only possible if a training sequence is received. Therefore, the estimator presented herein and those that assume the same model for the CCI can only operate in data-aided or decision-directed mode. The assumption that the signal shape is known should not be a too stringent one, since most communications or satellite navigation systems transmit certain training sequences and, subsequently the estimator can be switched to a decision-directed mode. In addition, in radar and sonar systems the shape of the received signal coincides with that of the transmitted one.

Some methods have focused on determining the time-delays of the multipath components together with some other parameters, for instance the directions of arrival (DOA), describing the channel [7, 8, 9]. Those methods exploit the full space-time structure of the multipath. Except for some cases that resort to a particular configuration of the antenna array, the primary drawback of these approaches is that complicated search procedures are required to estimate the desired parameters. Moreover, when DOA estimates are to be estimated, it is necessary to have a calibrated antenna array, which is a restrictive assumption. For these reasons, we will use an unstructured model for the channel. Although this leads to an increase in the number of unknowns with respect to a parameterized model, the dependence on the channel is linear and it can be estimated in closed form, as in [3].

A number of techniques that estimate the delays of each of the received replicas and assume that the value of each delay is arbitrary have been developed. However, receivers usually combine the different rays of the received signal using a RAKE structure. This structure can be viewed as a

\* Partial support provided by the Catalan and Spanish Governments under grants 1997FI 00755 APDT, TIC98-0412, TIC98-0703, TIC99-0849, and CIRIT 1998SGR-00081.

† Partial support provided by U.S. National Science Foundation Wireless Initiative Grant CCR 99-79452, and by the U.S. Office of Naval Research under Grant N00014-99-1-0692.



bank of filters, with a fixed delay (typically the inverse of the bandwidth or a fraction thereof) between each pair of filters. Hence, it seems logical to extend this structure also to the timing synchronization. The method proposed in this paper is based on the estimator in [4, 5]. The difference lies in that now a fixed separation between the received replicas is assumed, thereby only the *absolute* delay of the whole set of replicas has to be estimated. The interesting consequence of this model is that the delay can be obtained by finding the roots of a low-order polynomial, for which computationally efficient algorithms exist. Consequently, the method is specially tailored to applications where fast (feedforward) synchronization of the received signal is needed [2].

## 2. DATA MODEL

The signal received by an arbitrary  $m$  element array is modeled as

$$\mathbf{y}[n] = \sum_{k=1}^d \mathbf{h}_k s(nT_s - \tau - (k-1)T_0) + \mathbf{e}[n] \quad (1)$$

where  $T_s$  is the sampling period,  $\{\mathbf{h}_k\}$  are the FIR channel coefficient vectors and  $d$  is the temporal length of the channel.  $T_0$  is the temporal spacing of the FIR channel and can be freely chosen, together with  $d$ , when setting up the model (1). The transmitted signal  $s(t)$  is assumed to be known to within the scalar time delay parameter  $\tau$ . If  $N$  samples are collected, they all may be grouped together as follows:

$$\mathbf{Y} = [\mathbf{y}[1] \quad \mathbf{y}[2] \quad \dots \quad \mathbf{y}[N]] = \mathbf{H}\mathbf{S}(\tau) + \mathbf{E} \quad (2)$$

where  $\mathbf{E}$  is formed identically to  $\mathbf{Y}$  and

$$\mathbf{H} = [\mathbf{h}_1 \quad \mathbf{h}_2 \quad \dots \quad \mathbf{h}_d] \quad (3)$$

The  $m \times d$  matrix  $\mathbf{H}$  represents the single-input-multiple-output (SIMO) channel for the signal of interest. The  $p, q$ -th element of the matrix  $\mathbf{S}(\tau)$  is  $s(qT_s - \tau - (p-1)T_0)$ . The term  $\mathbf{e}[n]$ , which gathers the noise and all other CCI, is modeled as a complex, circularly-symmetric, zero-mean Gaussian process. It is assumed to be temporally white and spatially colored with an arbitrary unknown correlation matrix:

$$\mathcal{E}\{\mathbf{e}[n] \mathbf{e}^*[m]\} = \mathbf{Q} \delta_{n,m} \quad (4)$$

where  $(\cdot)^*$  denotes the complex conjugate transpose operation. While such a model for  $\mathbf{e}[n]$  is clearly only approximate, it captures the most significant effects of the noise and interference, and leads to tractable algorithms. For the asymptotic results in the next section to be valid, the following additional assumption is needed:  $s(t)$  is a band-limited finite-average-power signal, and the sampling period  $T_s$  satisfies the Nyquist criterion.

Note that though the pulse shaping filter could be factored into the channel matrix, as in [8], we assume herein that the elements of  $\mathbf{S}(\tau)$  are samples of the continuous modulated waveform  $s(t)$ . As such, the matrix  $\mathbf{H}$  only describes the propagation effects of the channel, and  $\tau$  is a continuous-valued parameter. This modeling premise is

different to those usually taken in other work addressing the equalization of FIR channels rather than the synchronization.

The model in (1) is closely related to that employed in other methods that attempt to estimate the delays of the different arrivals, such as [7, 6, 4, 5]. In those cases, the received vector model consists of the contribution of  $L$  arrivals as follows

$$\mathbf{Y} = \mathbf{A}\tilde{\mathbf{S}}(\tau) + \mathbf{E} \quad (5)$$

where

$$\boldsymbol{\tau} = [\tau_1 \quad \dots \quad \tau_L]^T \quad \mathbf{A} = [\mathbf{a}_1 \quad \dots \quad \mathbf{a}_L] \quad (6)$$

$$\mathbf{s}(\tau_l) = [s(T_s - \tau_l) \quad \dots \quad s(NT_s - \tau_l)] \quad (7)$$

$$\tilde{\mathbf{S}}(\tau) = [\mathbf{s}^T(\tau_1) \quad \dots \quad \mathbf{s}^T(\tau_L)]^T. \quad (8)$$

The columns of the matrix  $\mathbf{A}$  are the spatial signatures of the different arrivals. Assuming that the signal  $s(t)$  is band-limited and  $T_0$  satisfies the Nyquist criterion, each row of the matrix  $\tilde{\mathbf{S}}(\tau)$  can be expressed as a linear combination of the elements of  $\mathbf{S}(\tau)$  [10]. Therefore, there exist a  $L \times d$  interpolating matrix  $\mathbf{T}$  that satisfies

$$\tilde{\mathbf{S}}(\tau) = \mathbf{T}\mathbf{S}(\tau). \quad (9)$$

For the equality (9) to be exact in a general case, the column dimension of  $\mathbf{T}$ , that is  $d$ , should be infinite. However, very good approximations can be obtained for finite  $d$  [10]. Finally, identifying the channel matrix as  $\mathbf{H} = \mathbf{A}\mathbf{T}$ , the relationship between the models in (1) and (5) becomes apparent.

## 3. MAXIMUM LIKELIHOOD ESTIMATOR AND ASYMPTOTICALLY EQUIVALENT APPROXIMATION

Under the model described above, the negative log-likelihood function of the data  $\mathbf{Y}$  in (2) is given by (to within irrelevant constants)<sup>1</sup>

$$\Lambda(\tau, \mathbf{H}, \mathbf{Q}) = \log |\mathbf{Q}| + \text{Tr}\{\mathbf{C}(\tau, \mathbf{H}) \mathbf{Q}^{-1}\}, \quad (10)$$

where

$$\mathbf{C}(\tau, \mathbf{H}) = \hat{\mathbf{R}}_{yy} - \mathbf{H}\hat{\mathbf{R}}_{ys}^*(\tau) - \hat{\mathbf{R}}_{ys}(\tau)\mathbf{H}^* + \mathbf{H}\hat{\mathbf{R}}_{ss}(\tau)\mathbf{H}^* \quad (11)$$

$$\hat{\mathbf{R}}_{yy} = \frac{1}{N} \mathbf{Y} \mathbf{Y}^* \quad \hat{\mathbf{R}}_{ys}(\tau) = \frac{1}{N} \mathbf{Y} \mathbf{S}^*(\tau) \quad (12)$$

Since  $\mathbf{H}$  and  $\mathbf{Q}$  are taken as unstructured deterministic matrices, the minimization of (10) may be performed explicitly with respect to them. Their ML estimates may be expressed as

$$\hat{\mathbf{H}}_{ML}(\tau) = \hat{\mathbf{R}}_{ys}(\tau) \hat{\mathbf{R}}_{ss}^{-1}(\tau) \quad (13)$$

$$\hat{\mathbf{Q}}_{ML}(\tau) = \hat{\mathbf{R}}_{yy} - \hat{\mathbf{R}}_{ys}(\tau) \hat{\mathbf{R}}_{ss}^{-1}(\tau) \hat{\mathbf{R}}_{ys}^*(\tau). \quad (14)$$

<sup>1</sup>  $|\cdot|$  and  $\text{Tr}\{\cdot\}$  denote the determinant and the trace of a matrix, respectively.

Ignoring parameter independent constants, the resulting ML criterion for  $\tau$  is the minimizing argument of

$$f(\tau) = \log |\mathbf{I} - \mathbf{B}(\tau)| \quad (15)$$

where

$$\mathbf{B}(\tau) = \frac{1}{N} \hat{\mathbf{R}}_{yy}^{-1/2} \mathbf{Y} \mathbf{P}_{\mathbf{S}^*} \mathbf{Y}^* \hat{\mathbf{R}}_{yy}^{-1/2} \quad (16)$$

$$\mathbf{P}_{\mathbf{S}^*}(\tau) = \mathbf{S}^*(\tau) (\mathbf{S}(\tau) \mathbf{S}^*(\tau))^{-1} \mathbf{S}(\tau) \quad (17)$$

If the noise had been assumed spatially white (*i.e.*  $\mathbf{Q} = \sigma^2 \mathbf{I}$ ), the ML cost function, in place of (15), would have been

$$f^w(\tau) = -\frac{1}{N} \text{Tr} \{ \mathbf{Y} \mathbf{P}_{\mathbf{S}^*}(\tau) \mathbf{Y}^* \} \quad (18)$$

Since the dependence of this cost function on the projection matrix  $\mathbf{P}_{\mathbf{S}^*}$  is linear, the algorithm presented in the next section could be applied in order to find the minimum of (18) by rooting a polynomial. This interesting algorithm is not directly applicable to (15) because of the determinant operator. However, based on the results of [4, 5], it is straightforward to build a cost function that is linear in the projector  $\mathbf{P}_{\mathbf{S}^*}$  and yields asymptotically (large  $N$ , throughout the paper) equivalent estimates to those provided by  $f(\tau)$ . This alternative cost function to be minimized is

$$g(\tau, \hat{\mathbf{W}}) = -\text{Tr} \{ \hat{\mathbf{W}} \mathbf{B}(\tau) \} \quad (19)$$

The weighting matrix  $\hat{\mathbf{W}}$  is computed as

$$\hat{\mathbf{W}} = (\mathbf{I} - \mathbf{B}(\hat{\tau}))^{-1} \quad (20)$$

where  $\hat{\tau}$  is a consistent estimate of the true delay. This initial consistent estimate, that is when a previous estimate is not available to compute the weighting matrix, is simply obtained as the minimizing argument of  $g(\tau, \mathbf{I})$ . Following the development in [4, 5], it can be shown that both (15) and (19) provide, under mild conditions and in absence of modeling errors, consistent and asymptotically efficient estimates. Note that it can be argued that since  $N$  is the length of the training sequence, we will never reach asymptotics in  $N$ . However, the discussion above is completely meaningful because the numerical results show that the asymptotic behaviour is reached for rather modest sample sizes. It is not difficult to show that the CRB for the problem at hand is

$$\text{CRB}^{-1}(\tau) = 2 \text{Tr} \left\{ \left( \tilde{\mathbf{D}}(\tau) \mathbf{P}_{\tilde{\mathbf{S}}^*}^{\perp} \tilde{\mathbf{D}}^*(\tau) \right) (\mathbf{H}^* \mathbf{Q}^{-1} \mathbf{H}) \right\},$$

where the matrices  $\tilde{\mathbf{S}}$  and  $\tilde{\mathbf{D}}$  (which is the derivative of the former) are evaluated at

$$\tau = [\tau \quad \tau + T_0 \quad \dots \quad \tau + (d-1)T_0]^T.$$

#### 4. POLYNOMIAL ROOTING APPROACH

At this point we are concerned with the minimization of the following general expression

$$g(\tau, \mathbf{W}) = -\frac{1}{N} \text{Tr} \left\{ \mathbf{W}^{1/2} \hat{\mathbf{R}}_{yy}^{-1/2} \mathbf{Y} \mathbf{P}_{\mathbf{S}^*}(\tau) \mathbf{Y}^* \hat{\mathbf{R}}_{yy}^{-1/2} \mathbf{W}^{1/2} \right\} \quad (21)$$

For appropriate choices of  $\mathbf{W}$ , this expression represents the asymptotically efficient and the consistent estimators for correlated noise ( $g(\tau, \hat{\mathbf{W}})$  and  $g(\tau, \mathbf{I})$ ), and the white-noise ML estimator ( $f^w(\tau)$ ). Now, the  $N$  temporal samples are transformed into the frequency domain using the DFT, so that the signal approximately satisfies the following relationship<sup>2</sup>

$$\mathbf{S}^*(\tau) = \mathbf{S}_\omega^* \mathbf{V}(\tau) \quad (22)$$

where  $\mathbf{S}_\omega$  is a diagonal matrix whose entries are the DFT of the samples  $[s(T_s), \dots, s(N T_s)]$ , and

$$\mathbf{V}(\tau) = [\mathbf{v}(\tau) \quad \mathbf{v}(\tau + T_0) \quad \dots \quad \mathbf{v}(\tau + (d-1)T_0)]$$

$$\mathbf{v}(\tau) = [\exp(j\omega_1 \tau) \quad \dots \quad \exp(j\omega_N \tau)]^T \quad (23)$$

$$\omega_i = \frac{2\pi}{N T_s} \left( i - 1 - \text{floor} \left( \frac{N}{2} \right) \right) \quad (24)$$

The criterion in (21) may be expressed as a function of  $x \triangleq \exp(j2\pi\tau/NT_s)$ , resulting in a polynomial in  $x$  of order  $2N - 2$ , since  $(\mathbf{V}^*(\tau) \mathbf{S}_\omega \mathbf{S}_\omega^* \mathbf{V}(\tau))$  does not depend on  $x$ . This approach lacks of interest because  $N$  is generally large. Below we describe a method that leads to the rooting of polynomials of order  $2d$ , and it is natural that  $d \ll N$ .

Let the elements of the vector  $\mathbf{p} = [p_0 \dots p_d]^T$  be taken from the coefficients of the polynomial

$$p(z) = p_0 z^d + p_1 z^{d-1} + \dots + p_d \quad (25)$$

whose roots are:

$$\{r^0 x, r^1 x, \dots, r^{(d-1)} x\}, \quad (26)$$

where  $r \triangleq \exp(j2\pi T_0/NT_s)$ . If we define

$$\mathbf{F} = \mathbf{S}_\omega^{-*} \mathbf{Y}^* \hat{\mathbf{R}}_{yy}^{-1/2} \mathbf{W}^{1/2} / \sqrt{N} \quad (27)$$

$$\mathbf{\Psi} = (\mathbf{P}^* \mathbf{S}_\omega^{-*} \mathbf{S}_\omega^{-1} \mathbf{P})^{-1} \quad (28)$$

and build the  $N \times N - d$  matrix

$$\mathbf{P} = \begin{bmatrix} p_d & p_{d-1} & \dots & p_0 & 0 & 0 \\ 0 & p_d & p_{d-1} & \dots & p_0 & 0 \\ & \ddots & \ddots & \ddots & \ddots & \ddots \\ & & 0 & p_d & p_{d-1} & \dots & p_0 \end{bmatrix}^*, \quad (29)$$

then minimizing (21) is equivalent to minimizing [6]

$$\tilde{g}(\mathbf{p}, \mathbf{W}) = \text{Tr} \{ \mathbf{F}^* \mathbf{P} \mathbf{\Psi} \mathbf{P}^* \mathbf{F} \} \quad (30)$$

It can be readily shown that the vector  $\mathbf{p}$  satisfies

$$\mathbf{p} = \mathbf{K} \mathbf{t}(x) \quad (31)$$

where  $\mathbf{K}$  is diagonal matrix whose elements are the coefficients of the polynomial  $p(z)$  for the case  $x = 1$ , and

$$\mathbf{t}(x) = [1 \quad x \quad \dots \quad x^d]^T \quad (32)$$

<sup>2</sup>The same notation is used for both the time and frequency domains because the DFT is an unitary transformation and, therefore, the estimators presented in section 3 are identically applicable in the frequency domain.

Therefore, if the term  $\Psi$  is held fixed, the cost function in (30) can be written as a polynomial in  $x$  of order  $2d$ , as follows:

$$\tilde{g}(x, \mathbf{W}) = \mathbf{t}^T (1/x) \mathbf{K}^* \mathbf{C} \mathbf{K} \mathbf{t}(x) \quad (33)$$

for some matrix  $\mathbf{C}$  obtained from  $\mathbf{F}$  and  $\Psi$ . For sake of the brevity the explicit form of  $\mathbf{C}$  is omitted. The minimum of  $\tilde{g}(x, \mathbf{W})$  on the unit circle is computed by first finding the roots of its derivative. Next, (33) is evaluated at the set of roots that lie on the unit circle, and the one giving the minimum is selected. Using this root and the definition of  $x$ , the delay estimate is easily obtained. This procedure is repeated until convergence or failure conditions are satisfied (e.g., in the simulations these conditions are: change in  $x$  smaller than  $10^{-4}$ , number iterations larger than 50). At each iteration, the matrix  $\Psi$  is recomputed using the previous estimate of  $x$ ; and in the first iteration,  $\Psi$  is taken equal to the identity.

An essential feature of this algorithm is that the inverse matrix operation needed in the computation of  $\Psi$  needs to be calculated only once, and it can be done *off-line* since the matrix to be inverted depends exclusively on some design parameters. This follows from the fact that  $\Psi$  can be decomposed as

$$\Psi = \mathbf{U} \Psi_1 \mathbf{U}^{-1} \quad (34)$$

where  $\Psi_1$  is the value of  $\Psi$  for  $x = 1$ , and

$$\mathbf{U} = \text{diag} \{1, x, \dots, x^{N-d+1}\}. \quad (35)$$

Therefore, the update of  $\Psi$  at every iteration only involves the left and right-hand product of a fixed matrix by a diagonal one that solely depends on  $x$ .

## 5. NUMERICAL RESULTS

We analyze the performance of the estimators proposed in this paper, and compare it with the Cramér-Rao Bound (CRB). Specifically, we consider the exact ML estimator for the colored-noise case given by (15), its approximation in (19) and the ML estimator for the white-noise case in (18). The cost function of the first one is minimized by means of a search. Whereas, the polynomial rooting algorithm in section 4 is applied to the latter two. In the case of the approximate ML estimator, we have chosen to update the matrix  $\tilde{\mathbf{W}}$ , together with  $\Psi$ , at each iteration of the algorithm. The RMSE (root mean square error) are computed from 500 Monte Carlo realizations.

We concentrate on a scenario where  $L = 2$  delayed arrivals of a known signal are received by a uniform linear array with 6 antennas spaced  $0.5\lambda$  apart. This known signal is a concatenation of  $K$  truncated and sampled Nyquist square root raised cosine pulses. Each pulse has a bandwidth equal to  $(1 + \alpha)/2T_c$ , is truncated to the interval  $[-5T_c, 5T_c]$ , and the sampling period is  $T_s = T_c/2$ , so there are 21 samples in each pulse. The roll-off factor is set equal to  $\alpha = 0.2$ . The use of this type of signal is of interest because each pulse may represent the output of the despreaders at every symbol period in a direct-sequence CDMA system. For simplicity, the spatial signatures of the two arrivals are

the array steering vectors for DOAs equal to  $0^\circ$  and  $10^\circ$  relative to the broadside. The noise plus interference field in which the array operates consists of: *i*) spatially and temporally white Gaussian noise, and *ii*) a temporally white Gaussian interference at DOA  $-30^\circ$ . The remaining scenario parameters, except when one of them is varied, are as follows:  $K = 4$  pulses; delays of the two arrivals equal to  $\tau_1 = 0$  and  $\tau_2 = 0.5T_c$ ; signal to noise ratio (SNR) of the first arrival: 14dB; Signal to Interference Ratio (SIR) of the first arrival: -7dB; the second signal is attenuated 3dB with respect the first, and they are in phase at the first sensor. The temporal spacing of the FIR channel is assumed to be  $T_0 = 0.5T_c$ .

In figure 1, the finite-sample and asymptotic performance in absence of model errors (i.e.,  $T_0 = \tau_1 - \tau_0$  and the length of the FIR filter  $d$  is equal to the number of arrivals) of the different estimator is illustrated. We consider that the number of taps of the channel is  $d = 2$ . The RMSEs of the exact ML estimator and the proposed approximation reach the CRB for small sample sizes. This fact proves that neither the approximation leading to (19) nor the subsequent minimization using the polynomial rooting algorithm entail a significant degradation with respect to the exact search-based estimator. Figure 2 bears out that the methods that take into account the spatial correlation of the interference are practically insensitive to the CCI level, whenever enough degrees of freedom are available. On the other hand, under the rather usual of assumption of white-noise, the resulting estimator completely fails for  $\text{SIR} < -10\text{dB}$ . In figure 3, we investigate the performance of the estimators when for  $d = 4$  and the delay difference between the signal arrivals,  $\tau_2 - \tau_1$ , does not necessarily coincide with the spacing of the FIR channel,  $T_0$ . As expected, the RMSE presents minima when the former is a multiple of the latter. In the other cases, the model in (1) is only approximate, which results in a higher RMSE. Finally, increasing the length of the FIR filter beyond the necessary minimum ( $d = 2$  in this case) impairs the performance, as shown in figure 4.

## 6. CONCLUSIONS

The problem of time delay estimation in a multipath channel has been considered. The channel is modeled as an unknown FIR filter, and the CCI is assumed to have unknown spatial correlation. Starting from the exact ML solution, we have derived an approximate estimator, which has allowed us to use a polynomial rooting approach to obtain the estimates. The proposed method attains the CRB in absence of modeling errors and is robust against arbitrarily high interference levels. Finally, the effects of varying the number of taps of the channel and varying the delay between the arrivals of the signal have been investigated.

## REFERENCES

- [1] S. Parkvall, E. Ström, and B. Ottersten, "The Impact of Timing Errors on the Performance of Linear DS-SS Receivers," *IEEE Journal on Selected Areas on Communications*, vol. 14, pp. 1660-1668, Oct. 1996.

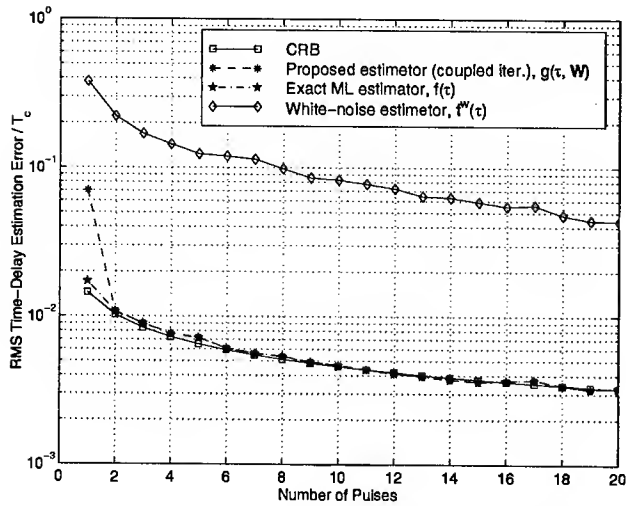


Figure 1: RMSE versus the number of training pulses.

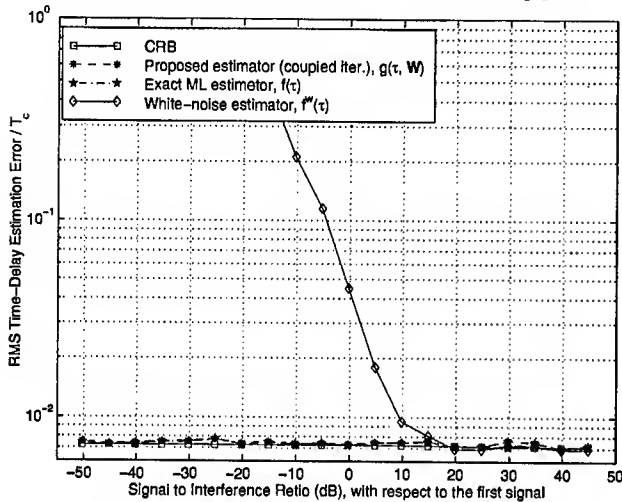


Figure 2: RMSE versus the interference power.

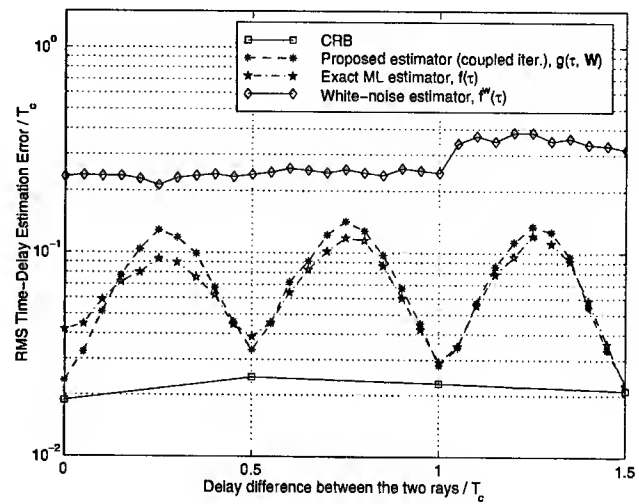


Figure 3: RMSE versus delay separation of the two arrivals.

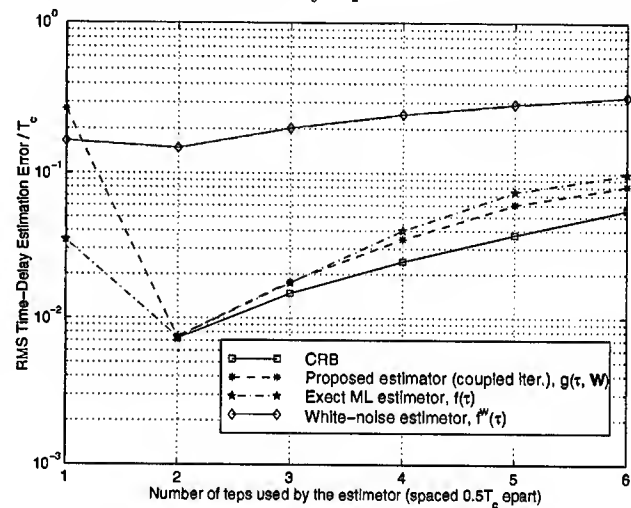


Figure 4: RMSE versus number of taps of the channel ( $d$ ).

- [2] M. Mengali and A. N. D'Andrea, *Synchronization Techniques for Digital Receivers*. Plenum Press, 1997.
- [3] D. Astély, A. Jakobsson, and A. Swindlehurst, "Burst Synchronization on Unknown Frequency Selective Channels with Co-Channel Interference Using an Antenna Array," in *Proc. IEEE Vehicular Technology Conf.*, (Houston, TX), March 1999.
- [4] G. Seco, A. Swindlehurst, J. A. Fernández-Rubio, and D. Astély, "A Reduced-Complexity and Asymptotically Efficient Time-Delay Estimator," in *Proc. ICASSP*, (Istanbul, Turkey), 2000.
- [5] G. Seco, A. L. Swindlehurst, and D. Astély, "Exploiting Antenna Arrays for Synchronization," in *Signal Processing Advances in Wireless Communications* (G. B. Giannakis, Y. Hua, P. Stoica, and L. Tong, eds.), vol. II: Trends in Single- and Multi-User Systems, ch. 10, Prentice-Hall, 2000. To be published.
- [6] A. L. Swindlehurst, "Time Delay and Spatial Signa-

ture Estimation Using Known Asynchronous Signals," *IEEE Trans. SP*, vol. 46, pp. 449-462, Feb. 1998.

- [7] M. Wax and A. Leshem, "Joint Estimation of Time Delays and Directions of Arrival of Multiple Reflections of a Known Signal," *IEEE Trans. on SP*, vol. 45, pp. 2477-2484, Oct. 1997.
- [8] G. Raleigh and T. Boros, "Joint Space-Time Parameter Estimation for Wireless Communication Channels," *IEEE Trans. on SP*, vol. 46, pp. 1333-1343, May 1998.
- [9] B. Fleury, M. Tshudin, R. Heddergott, D. Dahlhaus, and K. Pedersen, "Channel Parameter Estimation in Mobile Radio Environments Using the SAGE Algorithm," *IEEE J. Select. Areas Commun.*, vol. 17, pp. 434-450, March 1999.
- [10] T. I. Laakso, V. Välimäki, M. Karjalainen, and U. K. Laine, "Splitting the Unit Delay," *IEEE Signal Processing Magazine*, vol. 13, pp. 30-60, Jan. 1996.

# SUPER-EXPONENTIAL-ESTIMATOR FOR FAST BLIND CHANNEL IDENTIFICATION OF MOBILE RADIO FADING CHANNELS

*Andreas Schmidbauer*

Munich University of Technology (TUM)  
Institute for Communications Engineering (LNT)  
Arcisstraße 21, D-80290 Munich, Germany  
email: [Andreas.Schmidbauer@ei.tum.de](mailto:Andreas.Schmidbauer@ei.tum.de)

## ABSTRACT

An iterative algorithm for blind channel identification (no training symbols necessary) based on the Super-Exponential-Algorithm is shown. On the assumption of independent, identically distributed (i.i.d.) data the algorithm has fast convergence properties. It is robust with respect to system overfit (supernumerarily assumed channel coefficients converge to zero) and influence of modest additive white Gaussian noise even in mixed-phase moving average channels. Despite of the use of fourth order cumulants the complexity of the algorithm is rather low compared with alternative blind methodes. So the implementation on a signal processor (TMS320C40) was possible assuming GSM-like conditions.

## 1. INTRODUCTION

In recent years there were several suggestions for blind equalization, i.e., training sequences are not necessary for adapting the equalizer in the receiver. Only information on the modulation scheme and the statistics of the transmitted symbols is necessary. To obtain fast convergence despite of short block lengths a closed form solution based on fourth order cumulants is used. The Eigenvector Algorithm (EVA) [3] as well as the Super-Exponential Algorithm (Supex) [5] belong to this category. Applying a FIR structure, these algorithms approximate the MMSE (Minimum Mean Square Error) solution. Severe problems occur if zeros of the channel impulse response are on (or close to) the unit circle as it is likely in mobile communication environments.

To avoid these problems a Decision Feedback Equalizer [4], or a system of a channel estimator and Viterbi equalizer can be used. Combining the impulse response of the equalizer with the channel impulse response leads

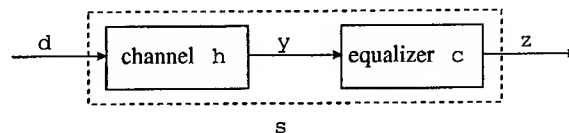


Figure 1: Time discrete system model of Supex Algorithm.

to an algorithm in closed form for channel identification based on the results of blind equalization. Thus, for example, the Eigenvector-Identification Algorithm (EVI) [1] is derived from EVA.

In the next section we describe the Supex algorithm for a linear equalizer. Based on that we derive in section 3 an algorithm for channel identification with comparable qualities as the EVI Algorithm but much lower complexity. Simulation results are given in section 4. A conclusion is given in the last section.

## 2. SUPER-EXPONENTIAL ALGORITHM (SUPEX) FOR A LINEAR EQUALIZER

In the time discrete system model as shown in Figure 1 the channel is represented by the column vector  $\mathbf{h}$ , whereas the column vector  $\mathbf{c}$  is used for the impulse response of the equalizer. The over-all system impulse response is represented by the column vector  $\mathbf{s}$ . The basic idea of the Supex Algorithm is to exponentiate iteratively each vector element of of the combined system  $\mathbf{s}$ , while holding the power of  $\mathbf{s}$  at a constant level. After a few iterations only the dominant amplitude element remains, while all other elements converge to zero. A normalization operation after each iteration

step leads to the following scheme:

$$\begin{aligned} \tilde{s}_n &= (s_n^{(i)})^p (s_n^{*(i)})^q \\ s_n^{(i+1)} &= \frac{\tilde{s}_n}{\|\tilde{s}\|} \end{aligned} \quad (1)$$

The expression  $(i)$  denotes the actual iteration step,  $*$  refers to the conjugate-transpose operation, and  $\|\dots\|$  stands for the Euclidean norm. The index  $n$  represents the  $n$ -th element of the vector  $\mathbf{s}$ . The auxiliary variable  $\tilde{s}_n$  is not really necessary but permits a foreseeable notation. As it is shown in [5] this algorithm converges to the desired dirac impulse (perfect equalization) if there is exactly one leading tap and  $p+q \geq 2$ . Unfortunately we don't know  $\mathbf{s}$ , since the channel  $\mathbf{h}$  is unknown to the receiver. That's why the algorithm (1) has to be expressed in terms of the equalizer coefficients  $\mathbf{c}$ . On the condition of transmitting i.i.d. samples  $\mathbf{d}$  and by choosing  $p = 2, q = 1$ , the channel  $\mathbf{h}$  can be expressed by using fourth order cumulants [5] and the autocorrelation matrix  $\mathbf{R}_{yy}$  of the received samples  $\mathbf{y}$ :

$$\begin{aligned} \tilde{\mathbf{c}} &= \sigma_d^2 \mathbf{R}_{yy}^{-1} \mathbf{v}_{zy}^{(i)} \\ \mathbf{c}^{(i+1)} &= \sigma_d \frac{\tilde{\mathbf{c}}}{\sqrt{\tilde{\mathbf{c}}^* \mathbf{R}_{yy} \tilde{\mathbf{c}}}} \end{aligned} \quad (2)$$

The vector  $\mathbf{v}_{zy}$  is calculated and updated after each iteration by the quotient

$$\mathbf{v}_{zy} = \frac{\text{cum}_4\{z_k, z_k, z_k^*, \mathbf{y}^*\}}{\text{cum}_4\{d_k, d_k, d_k^*, d_k^*\}} \quad (3)$$

of fourth order cumulants, where

$$\begin{aligned} \text{cum}_4\{x_1, x_2, x_3, x_4\} &= \text{E}\{x_1 x_2 x_3 x_4\} \\ &\quad - \text{E}\{x_1, x_2\} \text{E}\{x_3, x_4\} \\ &\quad - \text{E}\{x_1, x_3\} \text{E}\{x_2, x_4\} \\ &\quad - \text{E}\{x_1, x_4\} \text{E}\{x_2, x_3\}. \end{aligned} \quad (4)$$

The denominator in Eqn. (3) is constant because it depends only on the modulation scheme and the statistics of the transmitted data  $d_k$ .

The autocorrelation matrix  $\mathbf{R}_{yy}$  is given by

$$\mathbf{R}_{yy} = \text{E}\{\mathbf{y}\mathbf{y}^*\}, \quad (5)$$

with the column vector

$$\mathbf{y} = [y_k, y_{k-1}, \dots, y_{k-L}]' \quad (6)$$

of the last  $L$  received samples and  $L$  being the order (memory) of the equalizer.

The expression  $'$  denotes the transpose operation.

$\sigma_d^2$  comprises the power of the transmitted data and the output  $z_k$  of the equalizer is given at time  $t = k \cdot T$  by the convolution product  $z_k = y_k * c_k$ ,  $T$  being the symbol time.

Consecutive iteration steps are connected by  $z_k^{(i)} = y_k * c^{(i)}$  because  $\mathbf{v}_{zy}$  depends on  $z_k^{(i)}$  (compare equation (3)).

To start the iterations it is necessary to initialize the equalizer with

$$\mathbf{c}^{(0)} = [0, \dots, 0, 1, 0, \dots, 0]'. \quad (7)$$

If there is some knowledge about the channel  $\mathbf{h}$  (e.g. minimum, non-minimum phase) it is possible to set  $\mathbf{c}^{(0)}$  accordingly.

### 3. THE ALGORITHM FOR CHANNEL IDENTIFICATION: SUPEST

We already mentioned that the Supex Algorithm approximates the optimal MMSE solution. Hence, the result of the Supex Algorithm can be written as

$$\mathbf{c} \stackrel{!}{\approx} \mathbf{c}_{\text{MMSE}} = \mathbf{R}_{yy}^{-1} \cdot \mathbf{r}_{dy}. \quad (8)$$

The vector  $\mathbf{r}_{dy}$  is the crosscorrelation vector between the transmitted data and received samples. Of course this vector is unknown but assuming a linear modulation  $\mathbf{r}_{dy}$  can be expressed by:

$$\mathbf{r}_{dy} = \sigma_d^2 \cdot \mathbf{h}_r. \quad (9)$$

The vector  $\mathbf{h}_r$  contains the conjugate complex coefficients of the channel impulse response  $\mathbf{h}$  in reverse order (therefore  $r$ ). Applying Equations (8) and (9) we obtain:

$$\mathbf{c} \approx \mathbf{R}_{yy}^{-1} \cdot \sigma_d^2 \cdot \mathbf{h}_r. \quad (10)$$

Multiplying Equation (10) with  $\mathbf{R}_{yy}$  from the left-hand side and dividing by the power of the transmitted data leads to the estimated channel impulse response  $\tilde{\mathbf{h}}_r$ :

$$\tilde{\mathbf{h}}_r = \frac{1}{\sigma_d^2} \cdot \mathbf{R}_{yy} \cdot \mathbf{c}. \quad (11)$$

Using Equation (2) and reusing the scalar

$$K = \frac{\sigma_d}{\sqrt{\tilde{\mathbf{c}}^* \mathbf{R}_{yy} \tilde{\mathbf{c}}}} \quad (12)$$

which is already calculated during the last iteration of (2) we obtain the Equation of the Super-Exponential-Estimator (SupEst):

$$\tilde{\mathbf{h}}_r = K \cdot \mathbf{v}_{zy} \quad (13)$$

To achieve a proper adjustment of the vector  $\mathbf{v}_{zy}$ , several iteration steps of the Supex Algorithm are necessary before Equation (13) can be evaluated.

It is remarkable, assuming a channel with order  $L_h + 1$ , that  $2L_h + 1$  coefficients have to be estimated to make sure that all coefficients of a mixed phase channel are really in the result vector  $\tilde{\mathbf{h}}_r$ .

#### 4. RESULTS

We demonstrate the capabilities of the SupEst Algorithm with a channel of five taps as shown in Figure 2. There are two channel zeros close to the unit circle. Furthermore two coefficients of the impulse response are equal in amplitude violating the demand of one leading tap.

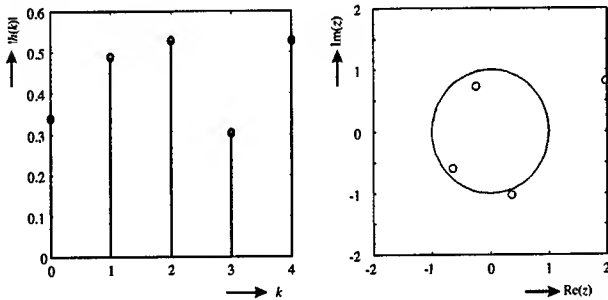


Figure 2: Absolute value of the amplitude impulse response and zero plot of MA channel.

Afterwards, we present the results of a measurement of a time variant mobile radio fading channel. The number of samples (130) was chosen even less than the block length of a GSM frame (148).

##### 4.1. Simulation with Matlab

Using QPSK modulation and coherent demodulation the simulation leads to the results depicted in Figure 3 (noise free case). Although there are only five taps according to the channel shown in Figure 2, we assumed a channel with nine coefficients. In other words,  $2 \cdot 8 + 1$  coefficients have to be estimated. Raising the block length the supernumerary coefficients converge to zero. The other coefficients converge to the original

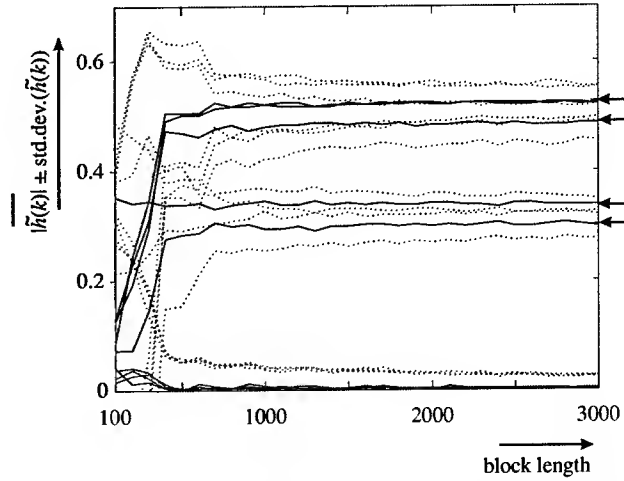


Figure 3: Results of SupEst Algorithm with increasing block length without noise.

taps marked by arrows on the right margins. With the dotted lines the standard deviation is marked.

In the next simulation, the influence of additive white Gaussian noise was investigated. Using the same conditions as above but with a fixed block length of 3000 samples, the results are shown in Figure 4. Up to an SNR of 6 dB the algorithm delivers good results.

These results remain valid for other channels and are even better in particular if there is only one leading tap.

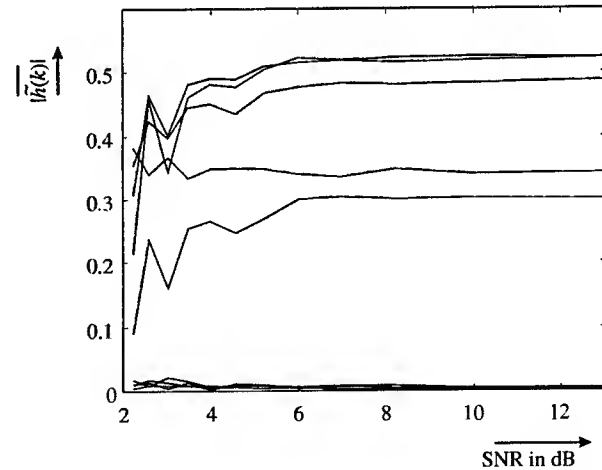


Figure 4: Behaviour of SupEst Algorithm with additive white Gaussian noise and a block length of 3000 samples.

## 4.2. Measurement

For the measurement a complete system similar to GSM (with a symbol rate of  $3.7 \cdot 10^{-6}$  s and a carrier frequency of 900 MHz) was used with BPSK modulation, transmission over a hardware fading channel emulator, and sampling. For this example, a three path channel (with path delays of a symbol time) was used. We assumed a speed of 10 km/h (for a clear figure, of course the channel identification for higher speeds is possible) and a block length of 130 samples. As can be seen in Figure 5, the SupEst Algorithm delivers the expected channel impulse response. Again, an overfit of the channel order did not affect the result.

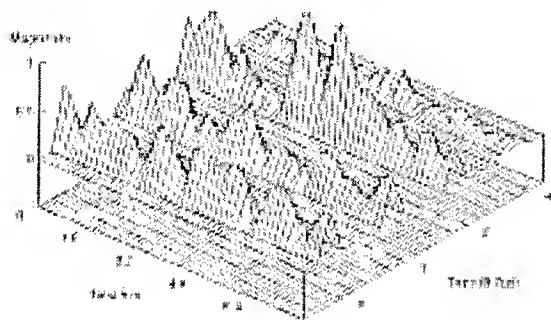


Figure 5: Absolute value of the estimated channel impulse response using the SupEst Algorithm for identification of a mobile radio fading channel with a block length of 130 samples.

## 5. CONCLUSION

We have shown that the new algorithm works with critical channels even if there are two taps with equal amplitudes. The SupEst is robust with respect to system overfit and influence of modest white Gaussian noise.

Although the algorithm uses fourth order cumulants, the computational effort is relatively low compared with other methods as the EVI Algorithm, for example. Therefore, we were able to implement the SupEst on a DSP in a mobile radio fading channel environment. As presented in Figure 5, the algorithm is performing well under conditions known from the GSM standard.

## REFERENCES

- [1] D. Boss, B. Jelonnek and K.-D. Kammeyer, "Eigenvector Algorithm for Blind MA System Identification," *EURASIP Signal Processing*, August 1995
- [2] J. Cadzow, "Blind deconvolution via cumulant extrema," *IEEE Signal Processing Magazine*, pp. 24 - 42, May 1996
- [3] B. Jelonnek and K.-D. Kammeyer, "A closed-form solution to blind equalization," *EURASIP Signal Processing*, Vol. 36, No. 3, pp. 251 - 259, 1994
- [4] A. Schmidbauer, M. Schmutz and R. Herzog, "Fast Blind Decision Feedback Equalizer for Mobile Links with Short Blocks," *VTC 1998*, Ottawa, pp. 2398 - 2402
- [5] O. Shalvi and E. Weinstein, "Super-Exponential Methods for blind deconvolution," *IEEE Transactions on Information Theory*, Vol. 39, No. 2, pp. 504 - 519, 1993



# FINITE DATA RECORD MAXIMUM SINR ADAPTIVE SPACE-TIME PROCESSING

*Ioannis N. Psaromiligkos and Stella N. Batalama*

Department of Electrical Engineering  
State University of New York at Buffalo  
Buffalo, NY 14260  
E-Mail: {ip2, batalama}@eng.buffalo.edu

## ABSTRACT

The presence of the desired signal during the estimation of the minimum-variance-distortionless-response (MVDR) or auxiliary-vector (AV) filter under limited data records leads to significant signal-to-interference-plus-noise ratio (SINR) performance degradation. We quantify this observation in the context of DS/CDMA communications by deriving two new close approximations for the probability density functions (under both desired-signal-“present” and “absent” conditions) of the output SINR and bit-error-rate (BER) of the sample-matrix-inversion (SMI) MVDR receiver. To avoid such performance degradation we propose a DS/CDMA receiver that utilizes a simple pilot-assisted algorithm that estimates and then subtracts the desired signal component from the received signal prior to filter estimation. Then, to accommodate decision directed operation we develop two recursive algorithms for the on-line estimation of the MVDR and AV filter and we study their convergence properties. Finally, simulation studies illustrate the BER performance of the overall receiver structure.

## 1. INTRODUCTION

The ideal minimum - variance - distortionless - response (MVDR) [1] filter evaluated using the perfectly known covariance matrix of the desired-signal-free input vector can be shown to be equivalent to the MVDR filter evaluated using the perfectly known signal-present covariance matrix. However, their estimated filter counterparts are not equivalent in terms of statistical performance measures of interest. In this paper we derive a close approximation of the probability density function (pdf) of the output signal-to-interference-plus-noise ratio (SINR) and the bit-error-rate (BER) of DS/CDMA receivers that utilize the following sample-matrix-inversion (SMI) MVDR filter estimates: The first estimate is calculated using desired-signal-free received vectors while the second estimate is calculated using desired-signal-present received vectors. The newly developed SINR and BER approximate pdf expressions prove and quantify the need for filter estimation (training) under desired-signal-free conditions (this need was also ob-

served long ago by array radar practitioners). In particular, comparing the two pdfs we will reason that the use of the desired-signal-present covariance matrix estimate can lead to significant SINR and BER performance degradation when the estimate is based on a limited record of input observations. To avoid the requirement for silent periods of the user of interest that requires significant coordination among the DS/CDMA users, we propose to proceed by estimating and then subtracting the desired transmission from the received vectors prior to the sample-average estimation of the covariance matrix. This way we obtain an estimate of the interference-plus-noise covariance matrix which is then used to evaluate the MVDR or the auxiliary-vector (AV) [2], [3] filter estimates of interest. To this end we propose to use initially a simple pilot-assisted (supervised) algorithm [4] and then switch to decision-directed mode. The latter operational mode, however, requires the use of on-line recursive algorithms for the estimation of the AV and MVDR filter. Recursive AV filter estimators have not been reported in the literature so far, while the LMS and RLS recursions qualify as candidates for the recursive on-line estimation of the MVDR filter. In this paper we develop a new recursive algorithm for the on-line estimation of the AV filter and a modified LMS-type algorithm for the estimation of the MVDR filter. These algorithms represent low-complexity alternatives to their batch counterparts. While their development was motivated by decision directed operation needs, they can be viewed as useful stand-alone tools, as well. Theoretical results included in this work establish formally the convergence of the proposed recursive algorithms. Finally, simulation studies illustrate the performance levels achieved by the overall proposed receiver structure that operates under limited pilot signaling followed by decision-directed mode.

## 2. SYSTEM MODEL AND BACKGROUND

We consider  $K$  DS/CDMA users that transmit over a multipath Rayleigh fading additive white Gaussian noise (AWGN) channel. The multipath channel is modeled as a tapped-delay line (TDL). The  $k$ -th user baseband transmitted signal is given by

$$u_k(t) = \sum_i b_k(i) \sqrt{E_k} s_k(t - iT), \quad k = 0, \dots, K-1, \quad (1)$$

THIS WORK WAS SUPPORTED IN PART BY THE NSF UNDER GRANT CCR-9805359 AND IN PART BY THE AFOSR UNDER GRANT F49620-99-1-0035.

where  $b_k(i) \in \{-1, +1\}$  is the  $i$ -th data (information) bit,  $T$  is the information bit period and  $E_k$  denotes the transmitted energy. The normalized signature waveform  $s_k(t)$  is given by  $s_k(t) = \sum_{l=0}^{L-1} d_k(l)\psi(t - lT_c)$  where  $d_k(l) \in \{-1, +1\}$  is the  $l$ -th bit of the spreading sequence of the  $k$ -th user,  $\psi(t)$  is the chip waveform,  $T_c$  is the chip period, and  $L = T/T_c$  is the system spreading gain.

The received signal is collected by a uniform linear antenna array consisting of  $M$  elements, spaced half-the-wavelength apart. The baseband received signal at the  $m$ -th antenna element ( $m = 0, \dots, M-1$ ) is given by

$$r_m(t) = \sum_{k=0}^{K-1} \sum_{n=0}^{N_p-1} c_{k,n} u_k(t - \frac{n}{B} - \tau_k) e^{-jm\pi \sin \theta_{k,n}} + n_m(t) \quad (2)$$

where  $N_p$  is the number of resolvable paths, assumed to be the same for all users. In (2),  $c_{k,n}$  is the effective complex Gaussian channel coefficient which is assumed to be identical across all antenna elements (no antenna diversity is considered) and  $\theta_{k,n}$  identifies the angle of arrival with respect to the  $n$ -th path of the  $k$ -th user.  $\tau_k$  is the relative transmission delay of user  $k$  with respect to user 0 ( $\tau_0 = 0$ ) and with  $r_m(t)$  bandlimited to  $B = 1/T_c$  the TDL has taps spaced at chip interval  $T_c$ . In (2),  $n_m(t)$  represents additive sensor noise modeled as temporally and spatially complex white Gaussian (WG) with variance  $\sigma^2$ . The received signals  $r_0(t), \dots, r_{M-1}(t)$  can be grouped to form the vector

$$\begin{aligned} \mathbf{r}(t) &\triangleq [r_0(t), r_1(t), \dots, r_{M-1}(t)]^T \\ &= \sum_{k=0}^{K-1} \sum_{n=0}^{N_p-1} c_{k,n} u_k(t - \frac{n}{B} - \tau_k) \mathbf{a}(\theta_k) + \mathbf{n}(t) \end{aligned} \quad (3)$$

where  $\mathbf{a}(\theta_k) \triangleq [1, e^{-j\pi \sin \theta_k}, \dots, e^{-j(M-1)\pi \sin \theta_k}]^T$ ,  $k = 0, \dots, K-1$ , is the steering vector associated with the  $k$ -th user, and  $\mathbf{n}(t) \triangleq [n_0(t), \dots, n_{M-1}(t)]^T$ .

Chip-matched filtering of  $\mathbf{r}(t)$  and sampling at the chip rate,  $1/T_c$ , over the symbol time interval ( $L + N_p - 1$  chip periods) prepares the data for one-shot detection of the  $i$ -th information bit of interest  $b_0(i)$ . By stacking the vector samples  $\mathbf{r}(0), \dots, \mathbf{r}((L + N_p - 1)T_c)$  one below the other we obtain the space-time received data vector

$$\bar{\mathbf{r}}_{M(L+N_p-1) \times 1} \triangleq [\mathbf{r}(0)^T \mathbf{r}(T_c)^T \dots \mathbf{r}((L+N_p-1)T_c)^T]^T \quad (4)$$

The cornerstone for any form of joint space-time filtering is the space-time signature which, for user 0, is defined as  $\mathbf{V}_0 \triangleq \sum_{n=0}^{N_p-1} \bar{\mathbf{S}}_0^{(n)} \otimes \mathbf{a}(\theta_{0,n})$  where

$$\bar{\mathbf{S}}_0^{(n)} \triangleq \underbrace{[0, \dots, 0]}_n, d_0(0), \dots, d_0(L-1), \underbrace{[0, \dots, 0]}_{N_p-n}^T \quad (5)$$

and  $\otimes$  denotes the Kronecker product. We assume (without loss of generality) that  $\|\mathbf{V}_0\| = 1$ .

A linear joint S-T receiver with tap weight vector  $\mathbf{w} \in \mathcal{C}^{M(L+N_p-1)}$  detects the transmitted bit of the user of interest as

$$\hat{b}_0 = \text{sgn}(\text{Re}\{\mathbf{w}^H \bar{\mathbf{r}}\}) \quad (6)$$

where  $\text{sgn}(\cdot)$  identifies the sign operation, and  $\text{Re}\{\cdot\}$  extracts the real part of a complex number. In this work we consider two types of linear receivers. The first type is the minimum-variance-distortionless-response (MVDR)

linear receiver whose tap-weight vector is designed to minimize the variance at the filter output  $E\{|\mathbf{w}^H \bar{\mathbf{r}}|^2\}$  while maintaining unity response in the vector direction  $\mathbf{V}_0$ . The MVDR-receiver tap weight vector is given by

$$\mathbf{w}_{MVDR} = \frac{\mathbf{R}^{-1} \mathbf{V}_0}{\mathbf{V}_0^H \mathbf{R}^{-1} \mathbf{V}_0} \quad (7)$$

In (7)  $\mathbf{R} \triangleq E\{\bar{\mathbf{r}}\bar{\mathbf{r}}^H\}$  is the covariance matrix of the received vector  $\bar{\mathbf{r}}$ . The second type is the Auxiliary-Vector (AV) linear receiver [2], [3] whose tap weight vector is a member of a sequence of vectors that converges to the MVDR solution. The AV filter sequence can be obtained as follows:

$$\mathbf{w}_{AV(0)} = \mathbf{V}_0 \quad (8)$$

$$\text{for } p = 1, 2, 3, \dots$$

$$\mathbf{G}_p = \mathbf{R} \mathbf{w}_{AV(p-1)} - \mathbf{V}_0^H \mathbf{R} \mathbf{w}_{AV(p-1)} \mathbf{V}_0 \quad (9)$$

$$\mu_p = \frac{\mathbf{G}_p^H \mathbf{R} \mathbf{w}_{AV(p-1)}}{\mathbf{G}_p^H \mathbf{R} \mathbf{G}_p} \quad (10)$$

$$\mathbf{w}_{AV(p)} = \mathbf{w}_{AV(0)} - \sum_{i=1}^p \mu_i \mathbf{G}_i \quad (11)$$

The auxiliary vector generation procedure may stop when  $\mathbf{G}_{p+1} = \mathbf{0}$ . In that case  $\mathbf{w}_{AV(p)}$  is exactly equal to  $\mathbf{w}_{MVDR}$ .

Formal theoretical analysis of the sequence of auxiliary-vector filters  $\mathbf{w}_{AV(0)}, \mathbf{w}_{AV(1)}, \dots$ , was pursued in [3] where it was shown that

$$\lim_{p \rightarrow \infty} \mathbf{w}_{AV(p)} = \mathbf{w}_{MVDR} \quad (12)$$

The MVDR and AV-type algorithms outlined above, require knowledge of the covariance matrix  $\mathbf{R}$  which is unknown in practice and it is usually estimated by sample averaging over a finite set of joint S-T data  $\bar{\mathbf{r}}_j$ ,  $j = 0, \dots, N-1$ .

1. The resulting estimator  $\hat{\mathbf{R}}$  is given by

$$\hat{\mathbf{R}} = \frac{1}{N} \sum_{j=0}^{N-1} \bar{\mathbf{r}}_j \bar{\mathbf{r}}_j^H \quad (13)$$

Using  $\hat{\mathbf{R}}$  in (7) and (8)-(11) we obtain the MVDR and AV filter estimates  $\hat{\mathbf{w}}_{SMI}$  and  $\hat{\mathbf{w}}_{AV(p)}$ , respectively, where the subscript SMI stands for "sample-matrix-inversion." As illustrated in [3] for a fixed finite data-record-size  $N$ , the sequence  $\{\hat{\mathbf{w}}_{AV(p)}\}_p$  provides filter estimators with varying bias versus covariance characteristics that converge to  $\hat{\mathbf{w}}_{SMI}$ . For short data records  $N$ , the early, non-asymptotic, elements of the generated sequence of AV estimators offer favorable bias/covariance balance and are seen to outperform significantly in mean-square estimation error the  $\hat{\mathbf{w}}_{SMI}$  estimator.

### 3. SINR AND BER PDFS OF THE JOINT S-T SMI MVDR RECEIVER

Let  $\mathbf{w}$  be a linear S-T receiver that is distortionless in the  $\mathbf{V}_0$  direction i.e.,  $\mathbf{w}^H \mathbf{V}_0 = 1$ . Then the SINR at the filter output is given by

$$S(\mathbf{w}) = \frac{E_0}{\mathbf{w}^H \mathbf{R}_{I+n} \mathbf{w}}, \quad (14)$$

where the index  $I+n$  is used to distinguish the interference-plus-noise input covariance matrix  $\mathbf{R}_{I+n} = \mathbf{R} - E_0 \mathbf{V}_0 \mathbf{V}_0^H$  from the desired-signal-present input covariance matrix  $\mathbf{R}$ .

We are interested in obtaining the pdf of the output SINR of the MVDR filter estimators  $\hat{\mathbf{w}}_{SMI} = \frac{\hat{\mathbf{R}}^{-1}\mathbf{v}_0}{\mathbf{v}_0^H \hat{\mathbf{R}}^{-1} \mathbf{v}_0}$  and  $\hat{\mathbf{w}}_{SMI,I+n} = \frac{\hat{\mathbf{R}}_{I+n}^{-1}\mathbf{v}_0}{\mathbf{v}_0^H \hat{\mathbf{R}}_{I+n}^{-1} \mathbf{v}_0}$  using  $N$ -point sample-average estimates of the desired-signal-present and desired-signal-free covariance matrix, respectively. Evaluation of these pdfs requires knowledge of the pdf of  $\hat{\mathbf{R}}$  and  $\hat{\mathbf{R}}_{I+n}$ . We make the following simplifying assumption: We assume that the received vectors are identically distributed according to a multivariate Gaussian distribution  $\mathcal{N}(\mathbf{0}, \mathbf{R})$ . Then, the estimator  $\hat{\mathbf{R}}$  is distributed according to a Wishart distribution with  $N$  degrees of freedom,  $\mathcal{W}_{M(L+N_p-1)}(N^{-1}\mathbf{R}; N)$  [5]. Similarly, we assume that  $\hat{\mathbf{R}}_{I+n}$  is distributed according to a Wishart distribution with  $N$  degrees of freedom,  $\mathcal{W}_{M(L+N_p-1)}(N^{-1}\mathbf{R}_{I+n}; N)$ .

The following theorem provides a close approximation of the pdf of the output SINR of the estimated SMI MVDR filter for the case in which filter estimation is performed in the presence of the desired signal as well as the case in which filter estimation is performed in the absence of the desired signal. The proof is omitted due to lack of space.

**Theorem 1** (i) The pdf of  $\mathcal{S}(\hat{\mathbf{w}}_{SMI})$  can be approximated by

$$f_S(s) = \frac{N(1+S_0)S_0[(N-M(L+N_p-1)+2)S_0 + 2\sqrt{2\pi}[Ns(S_0-s)(1+S_0)]^{3/2}}{s(M(L+N_p-1)(S_0+2) - N - 2S_0 - 4)]} e^{-\frac{[(M(L+N_p-1)-2)(1+s)S_0 - N(S_0-s)]^2}{2Ns(S_0-s)(1+S_0)}} \quad (15)$$

(ii) The pdf of  $\mathcal{S}(\hat{\mathbf{w}}_{SMI,I+n})$  can be approximated by

$$f_{S,I+n}(s) = \frac{NS_0[(2M(L+N_p-1) - N - 4)s + (N - M(L+N_p-1) + 2)S_0]}{2\sqrt{2\pi}(Ns(S_0-s))^{3/2}} e^{-\frac{(N-M(L+N_p-1)+2-N\frac{s}{S_0})^2}{2N\frac{s}{S_0}(1-\frac{s}{S_0})}} \quad (16)$$

where  $S_0 \triangleq \mathcal{S}(\mathbf{w}_{MVDR}) = \mathcal{S}(\mathbf{w}_{MVDR,I+n})$  is the output SINR of the ideal filters  $\mathbf{w}_{MVDR}$  and  $\mathbf{w}_{MVDR,I+n}$ ,  $N$  is the data record size,  $M$  is the number of antenna elements,  $L$  is the system processing gain, and  $N_p$  is the number of resolvable paths.  $\square$

In the following, we examine how the results of Theorem 1 which are based on the filter output SINR translate into BER terms. Under the assumption that the received vector is a Gaussian distributed the BER  $P_e(\mathbf{w})$  at the output of a sign detector that follows an arbitrary linear filter  $\mathbf{w}$  (distortionless in the  $\mathbf{V}_0$  direction) can be expressed as follows:

$$P_e(\mathbf{w}) \simeq Q\left(\sqrt{\mathcal{S}(\mathbf{w})}\right) \quad (17)$$

where  $\mathcal{S}(\cdot)$  is defined as in (14). The following proposition provides a close approximation of the pdf of the BER of the estimated MVDR receiver for the desired-signal-present and desired-signal-absent case. The proof is omitted due to lack of space.

**Proposition 1** (i) The pdf of  $P_e(\hat{\mathbf{w}}_{SMI})$  can be approximated by

$$f_{P_e}(x) = 2\sqrt{2\pi}Q^{-1}(x)f_S([Q^{-1}(x)]^2)e^{[Q^{-1}(x)]^2/2} \quad (18)$$

(ii) The pdf of  $P_e(\hat{\mathbf{w}}_{SMI,I+n})$  can be approximated by

$$f_{P_e,I+n}(x) = 2\sqrt{2\pi}Q^{-1}(x)f_{S,I+n}([Q^{-1}(x)]^2)e^{[Q^{-1}(x)]^2/2} \quad (19)$$

where  $f_S(s)$  and  $f_{S,I+n}(s)$  are given by (15) and (16), respectively.  $\square$

In Fig. 1 we plot the pdf of  $P_e(\hat{\mathbf{w}}_{SMI,I+n})$  and  $P_e(\hat{\mathbf{w}}_{SMI})$  for a DS/CDMA system for which the ideal MVDR performance level is at  $P_e = 10^{-2}$ . The data-record-size  $N$  is equal to 200. An antenna array consisting of  $M = 5$  elements is assumed. The system processing gain is  $L = 15$  and the number of paths is  $N_p = 3$ . Comparing the two pdfs, we see that the BER performance of  $\hat{\mathbf{w}}_{SMI,I+n}$  is significantly more likely to lie near the performance of the ideal filter ( $P_e(\mathbf{w}_{MVDR,I+n}) = P_e(\mathbf{w}_{MVDR}) = 10^{-2}$ ).

#### 4. INTERFERENCE-PLUS-NOISE COVARIANCE MATRIX ESTIMATION

The theoretical developments of the previous section reveal the advantages of evaluating the MVDR filter using an estimate of the interference-plus-noise covariance matrix  $\hat{\mathbf{R}}_{I+n}$  in place of the estimated covariance matrix  $\hat{\mathbf{R}}$ . Although the estimate  $\hat{\mathbf{R}}_{I+n}$  can be obtained easily during the silent periods of the user of interest, such an approach requires significant amount of coordination among users. Alternatively, we propose to estimate the interference-plus-noise component by subtracting an estimate of the desired transmission from the received samples.

The proposed algorithm is based on the observation that  $E\{b_0(j)\bar{\mathbf{r}}_j\} = \sqrt{E_0}\mathbf{V}_0$  [4]. Thus, given a known transmitted bit sequence  $\{b_0(j)\}_j$  we may estimate the product  $\sqrt{E_0}\mathbf{V}_0$  by  $\frac{1}{N}\sum_{i=1}^N b_0(i)\bar{\mathbf{r}}_i$  and the interference-plus-noise component of the vectors  $\bar{\mathbf{r}}_j$  by  $\tilde{\mathbf{r}}_j = \bar{\mathbf{r}}_j - b_0(j)\frac{1}{N}\sum_{i=1}^N b_0(i)\bar{\mathbf{r}}_i$ . The interference-plus-noise covariance matrix estimate is then given by  $\hat{\mathbf{R}}_{I+n} = \frac{1}{N}\sum_{j=1}^N \tilde{\mathbf{r}}_j \tilde{\mathbf{r}}_j^H$ . A recursive implementation of the  $\hat{\mathbf{R}}_{I+n}$  estimator is summarized below:

$$\mathbf{v}_0 = \mathbf{0} \quad (20)$$

$$\hat{\mathbf{R}}_{I+n}^{(0)} = \delta \mathbf{I}, \quad \delta > 0, \quad (21)$$

$$\text{for } j = 1, 2, \dots$$

$$\mathbf{v}_j = \frac{1}{j}((j-1)\mathbf{v}_{j-1} + b_0(j)\bar{\mathbf{r}}_j), \quad (22)$$

$$\tilde{\mathbf{r}}_j = \bar{\mathbf{r}}_j - b_0(j)\mathbf{v}_j, \quad (23)$$

$$\hat{\mathbf{R}}_{I+n}^{(j)} = \frac{1}{j}((j-1)\hat{\mathbf{R}}_{I+n}^{(j-1)} + \tilde{\mathbf{r}}_j \tilde{\mathbf{r}}_j^H), \quad (24)$$

The following theorem deals with the properties of the estimator  $\hat{\mathbf{R}}_{I+n}$  in (24). The proof is omitted due to lack of space.

**Theorem 2** The estimator  $\hat{\mathbf{R}}_{I+n}$  in (24) is an asymptotically unbiased estimator of  $\mathbf{R}_{I+n}$ . Moreover, for fixed  $j$ ,  $\hat{\mathbf{R}}_{I+n}^{(j)}$  is a biased estimator with bias

$$E\{\hat{\mathbf{R}}_{I+n}^{(j)}\} = \left[\frac{1}{j}\psi(j+1) + \gamma\right]\mathbf{R}_{I+n} \quad (25)$$

where  $\psi(\cdot)$  is the Digamma function and  $\gamma$  is the Euler constant.  $\square$

At the end of the training period the receiver reverts to decision-directed operation where the information bit  $b_0(j)$  in (22) and (23) is substituted by the estimate

$$\hat{b}_0(j) = \text{sgn} \left\{ \text{Re} \left[ \hat{\mathbf{w}}^{(j-1)H} \tilde{\mathbf{r}}_j \right] \right\}. \quad (26)$$

In (26),  $\hat{\mathbf{w}}^{(j-1)}$  is an MVDR or AV-type filter estimate evaluated using the covariance matrix estimate  $\hat{\mathbf{R}}_{I+n}^{(j-1)}$  of the previous  $j-1$  step. Implementation of the decision directed version of the algorithm in (20)-(24) is straightforward but computationally inefficient: At each step of the algorithm new filter estimates have to be formed based on the updated estimate of the covariance matrix  $\hat{\mathbf{R}}_{I+n}^{(j)}$ . In the next section we derive filter update rules that are based on the theory of stochastic approximation and provide simple, computationally efficient on-line recursions for the evaluation of the AV and MVDR filter.

## 5. RECURSIVE ON-LINE ESTIMATION OF THE AV AND MVDR FILTER

The single-AV case  $\mathbf{w}_{AV(1)} = \mathbf{V}_0 - \mu_1 \mathbf{G}_1$  can be treated as follows. We note that the auxiliary vector  $\mathbf{G}_1$  can be expressed in the form:

$$\mathbf{G}_1 = (\mathbf{I} - \mathbf{V}_0 \mathbf{V}_0^H) \mathbf{R} \mathbf{V}_0. \quad (27)$$

Equivalently, we may find  $\mathbf{G}_1$  as the unique solution of the equation  $Z'_1(\Gamma) = 0$  where

$$Z'_1(\Gamma) \triangleq \Gamma - (\mathbf{I} - \mathbf{V}_0 \mathbf{V}_0^H) \mathbf{R}_{I+n} \mathbf{V}_0. \quad (28)$$

On the other hand, the steering scalar  $\mu_1$  minimizes the mean square error between  $\mathbf{V}_0^H \tilde{\mathbf{r}}$  and  $\mathbf{G}_1^H \tilde{\mathbf{r}}$  and is given by (cf. (10))

$$\mu_1 = \frac{\mathbf{G}_1^H \mathbf{R}_{I+n} \mathbf{V}_0}{\mathbf{G}_1^H \mathbf{R}_{I+n} \mathbf{G}_1}. \quad (29)$$

Thus,  $\mu_1$  is the unique solution of the equation  $Z''_1(\nu) = 0$  with

$$Z''_1(\nu) \triangleq \nu (\mathbf{G}_1^H \mathbf{R}_{I+n} \mathbf{G}_1) - \mathbf{G}_1^H \mathbf{R}_{I+n} \mathbf{V}_0. \quad (30)$$

If we define the functions  $\zeta'_1(\Gamma; \tilde{\mathbf{r}}_j)$  and  $\zeta''_1(\nu; \tilde{\mathbf{r}}_j)$  as

$$\zeta'_1(\Gamma; \tilde{\mathbf{r}}_j) \triangleq \Gamma - \frac{j}{j-1} (\mathbf{I} - \mathbf{V}_0 \mathbf{V}_0^H) \tilde{\mathbf{r}}_j \tilde{\mathbf{r}}_j^H \mathbf{V}_0 \quad (31)$$

$$\zeta''_1(\nu; \tilde{\mathbf{r}}_j) \triangleq \nu \frac{j}{j-1} (\mathbf{G}_1^H \tilde{\mathbf{r}}_j \tilde{\mathbf{r}}_j^H \mathbf{G}_1) - \frac{j}{j-1} \mathbf{G}_1^H \tilde{\mathbf{r}}_j \tilde{\mathbf{r}}_j^H \mathbf{V}_0 \quad (32)$$

then  $E\{\zeta'_1(\Gamma; \tilde{\mathbf{r}}_j)\} = Z'_1(\Gamma)$  while  $E\{\zeta''_1(\nu; \tilde{\mathbf{r}}_j)\} = Z''_1(\nu)$ . The following theorem describes recursive procedures for the evaluation of  $\mathbf{G}_1$  and  $\mu_1$  and establishes their convergence w.p. 1 to the desired values. The proof is omitted due to lack of space.

**Theorem 3** Let  $\{\alpha_j\}_j$  be a sequence of positive numbers such that  $\sum_{j=1}^{+\infty} \alpha_j = \infty$  and  $\sum_{j=1}^{+\infty} \alpha_j^2 < \infty$ . The recursions

$$\hat{\mathbf{G}}_1^{(j)} = \hat{\mathbf{G}}_1^{(j-1)} + \alpha_j \zeta'_1(\hat{\mathbf{G}}_1^{(j-1)}; \tilde{\mathbf{r}}_j) \quad (33)$$

and

$$\hat{\mu}_1^{(j)} = \hat{\mu}_1^{(j-1)} + \alpha_j \zeta''_1(\hat{\mu}_1^{(j-1)}; \tilde{\mathbf{r}}_j) \quad (34)$$

converge w.p. 1 to  $\mathbf{G}_1$  and  $\mu_1$ , respectively, where  $\zeta'_1(\Gamma; \tilde{\mathbf{r}}_j)$  and  $\zeta''_1(\nu; \tilde{\mathbf{r}}_j)$  are defined by (31) and (32).  $\square$

In practice we evaluate  $\mathbf{G}_1$ , and  $\mu_1$  by coupling the two recursions of theorem 3 as follows. From (28) and (30) we see that the pair  $(\mathbf{G}_1, \mu_1)$  is the unique solution of the equation  $Z_1(\Gamma, \nu) = 0$  where the vector function  $Z_1(\Gamma, \nu)$  is defined as

$$Z_1(\Gamma, \nu) \triangleq \begin{bmatrix} \Gamma - (\mathbf{I} - \mathbf{V}_0 \mathbf{V}_0^H) \mathbf{R}_{I+n} \mathbf{V}_0 \\ \nu (\Gamma^H \mathbf{R}_{I+n} \Gamma) - \Gamma^H \mathbf{R}_{I+n} \mathbf{V}_0 \end{bmatrix} \quad (35)$$

If we define the vector function  $\zeta_1(\Gamma, \nu; \tilde{\mathbf{r}}_j)$  as

$$\zeta_1(\Gamma, \nu; \tilde{\mathbf{r}}_j) \triangleq \begin{bmatrix} \Gamma - (\mathbf{I} - \mathbf{V}_0 \mathbf{V}_0^H) \frac{j}{j-1} \tilde{\mathbf{r}}_j \tilde{\mathbf{r}}_j^H \mathbf{V}_0 \\ \nu (\Gamma^H \frac{j}{j-1} \tilde{\mathbf{r}}_j \tilde{\mathbf{r}}_j^H \Gamma) - \Gamma^H \frac{j}{j-1} \tilde{\mathbf{r}}_j \tilde{\mathbf{r}}_j^H \mathbf{V}_0 \end{bmatrix} \quad (36)$$

then we have

$$E\{\zeta_1(\Gamma, \nu; \tilde{\mathbf{r}}_j)\} = Z_1(\Gamma, \nu). \quad (37)$$

Thus we may use the recursion

$$[\hat{\mathbf{G}}_1^{(j)T}, \hat{\mu}_1^{(j)T}]^T = [\hat{\mathbf{G}}_1^{(j-1)T}, \hat{\mu}_1^{(j-1)T}]^T - \alpha_j \zeta_1(\hat{\mathbf{G}}_1^{(j-1)}, \hat{\mu}_1^{(j-1)}; \tilde{\mathbf{r}}_j) \quad (38)$$

to evaluate  $\mathbf{G}_1$  and  $\mu_1$ .

The extension to the multiple-AV case is straightforward. The  $m$ -th auxiliary vector  $\mathbf{G}_m$  and the  $m$ -th steering scalar  $\mu_m$  are the unique solutions with respect to  $\Gamma_m$  and  $\nu_m$ , respectively, of the equations

$$\Gamma_m - (\mathbf{I} - \mathbf{V}_0 \mathbf{V}_0^H) \mathbf{R}_{I+n} (\mathbf{V}_0 - \sum_{j=1}^{m-1} \mu_j \mathbf{G}_j) = 0 \quad (39)$$

and

$$\nu_m (\mathbf{G}_m^H \mathbf{R}_{I+n} \mathbf{G}_m) - \mathbf{G}_m^H \mathbf{R}_{I+n} (\mathbf{V}_0 - \sum_{j=1}^{m-1} \mu_j \mathbf{G}_j) = 0. \quad (40)$$

Thus, if we define the functions  $Z_m(\Gamma_m, \nu_m, \dots, \Gamma_1, \nu_1)$  as

$$Z_m(\Gamma_m, \nu_m, \dots, \Gamma_1, \nu_1) \triangleq \begin{bmatrix} \Gamma_m - (\mathbf{I} - \mathbf{V}_0 \mathbf{V}_0^H) \mathbf{R}_{I+n} (\mathbf{V}_0 - \sum_{j=1}^{m-1} \nu_j \Gamma_j) \\ \nu_m (\Gamma_m^H \mathbf{R}_{I+n} \Gamma_m) - \Gamma_m^H \mathbf{R}_{I+n} (\mathbf{V}_0 - \sum_{j=1}^{m-1} \nu_j \Gamma_j) \end{bmatrix} \quad (41)$$

then the  $M$  pairs  $(\mathbf{G}_1, \mu_1), \dots, (\mathbf{G}_M, \mu_M)$  are the unique solution of the system of equations

$$Z_1(\Gamma_1, \nu_1) = 0 \quad (42)$$

$\vdots$

$$Z_M(\Gamma_M, \nu_M, \dots, \Gamma_1, \nu_1) = 0. \quad (43)$$

Let  $\zeta_m(\Gamma_m, \nu_m, \dots, \Gamma_1, \nu_1; \tilde{\mathbf{r}}_j)$  be defined as

$$\zeta_m(\Gamma_m, \nu_m, \dots, \Gamma_1, \nu_1; \tilde{\mathbf{r}}_j) \triangleq \begin{bmatrix} \Gamma_m - \frac{j}{j-1} (\mathbf{I} - \mathbf{V}_0 \mathbf{V}_0^H) \tilde{\mathbf{r}}_j \tilde{\mathbf{r}}_j^H (\mathbf{V}_0 - \sum_{j=1}^{m-1} \nu_j \Gamma_j) \\ \frac{j}{j-1} \nu_m (\Gamma_m^H \tilde{\mathbf{r}}_j \tilde{\mathbf{r}}_j^H \Gamma_m) - \frac{j}{j-1} \Gamma_m^H \tilde{\mathbf{r}}_j \tilde{\mathbf{r}}_j^H (\mathbf{V}_0 - \sum_{j=1}^{m-1} \nu_j \Gamma_j) \end{bmatrix}. \quad (44)$$

Since  $E\{\zeta_m(\Gamma_m, \nu_m, \dots, \Gamma_1, \nu_1; \tilde{\mathbf{r}}_j)\} = Z_m(\Gamma_m, \nu_m, \dots, \Gamma_1, \nu_1)$  we use the following recursions

$$[\hat{\mathbf{G}}_1^{(j)T}, \hat{\mu}_1^{(j)T}]^T = [\hat{\mathbf{G}}_1^{(j-1)T}, \hat{\mu}_1^{(j-1)T}]^T - \alpha_j \zeta_1(\hat{\mathbf{G}}_1^{(j-1)}, \hat{\mu}_1^{(j-1)}; \tilde{\mathbf{r}}_j) \quad (45)$$

$$\vdots \quad (46)$$

$$[\hat{\mathbf{G}}_M^{(j)T}, \hat{\mu}_M^{(j)T}]^T = [\hat{\mathbf{G}}_M^{(j-1)T}, \hat{\mu}_M^{(j-1)T}]^T - \alpha_j \zeta_M(\hat{\mathbf{G}}_M^{(j-1)}, \hat{\mu}_M^{(j-1)}, \dots, \hat{\mathbf{G}}_1^{(j-1)}, \hat{\mu}_1^{(j-1)}; \tilde{\mathbf{r}}_j) \quad (47)$$

to evaluate the pairs  $(\mathbf{G}_1, \mu_1), \dots, (\mathbf{G}_M, \mu_M)$ . The initial values  $(\hat{\mathbf{G}}_1^{(0)}, \hat{\mu}_1^{(0)}), \dots, (\hat{\mathbf{G}}_M^{(0)}, \hat{\mu}_M^{(0)})$  are the corresponding values obtained at the end of the training period.

In Fig. 2a and 2b we examine the convergence of the proposed algorithm for the evaluation of the AV filter. A DS/CDMA system with  $K = 4$  users and processing gain  $L = 15$  is assumed. The number of resolvable paths is  $N_p = 3$  while the receiver's antenna array consists of  $M = 5$  elements. The users SNRs are fixed at 10, 13, 14, 15dB. The two (2) auxiliary vector filter case is considered. In Fig. 2a we plot the normalized cross-correlation between  $\hat{\mathbf{G}}_1^{(j)}, \hat{\mathbf{G}}_2^{(j)}$  and their ideal counterparts  $\mathbf{G}_1, \mathbf{G}_2$ , respectively, as a function of the iteration  $j$ . In Fig 2b we plot the mean absolute error between  $\hat{\mu}_1^{(j)}, \hat{\mu}_2^{(j)}$  and their ideal counterparts  $\mu_1, \mu_2$ , respectively, as a function of the iteration  $j$ . The results presented are averages over 500 independent experiments. The angles of arrival and delays of all the users are chosen randomly and kept constant for 5 experiments.

Following a similar reasoning we can derive a recursive stochastic algorithm for the evaluation of the MVDR filter. We recall that the MVDR filter evaluated in the absence of the desired signal is given by  $\mathbf{w}_{MVDR, I+n} = \frac{\mathbf{R}_{I+n}^{-1} \mathbf{V}_0}{\mathbf{V}_0^H \mathbf{R}_{I+n}^{-1} \mathbf{V}_0}$  and is equivalent in the SINR and BER sense to the filter  $\mathbf{w}'_{MVDR, I+n} = \mathbf{R}_{I+n}^{-1} \mathbf{V}_0$ . The latter filter is the unique solution of the equation  $\mathbf{R}_{I+n} \mathbf{w}'_{MVDR, I+n} - \mathbf{V}_0 = \mathbf{0}$ . If we define the function  $\zeta(\mathbf{w}; \tilde{\mathbf{r}}_j)$  as

$$\zeta(\mathbf{w}; \tilde{\mathbf{r}}_j) \triangleq \frac{j}{j-1} (\tilde{\mathbf{r}}^H \mathbf{w}) \tilde{\mathbf{r}} - \mathbf{V}_0 \quad (48)$$

then  $E\{\zeta(\mathbf{w}; \tilde{\mathbf{r}}_j)\} = \mathbf{R}_{I+n} \mathbf{w} - \mathbf{V}_0$ .

Thus, to evaluate  $\mathbf{w}'_{MVDR, I+n}$  it suffices to find the filter that makes the expected value of  $\zeta(\mathbf{w}; \tilde{\mathbf{r}}_j)$  equal to zero. The following proposition describes a recursive on-line stochastic procedure for the evaluation of  $\mathbf{w}'_{MVDR, I+n}$  and proves its convergence. The proof is omitted due to lack of space.

**Proposition 2** The recursion

$$\hat{\mathbf{w}}_{MVDR, I+n}^{(j)} = \hat{\mathbf{w}}_{MVDR, I+n}^{(j-1)} - \alpha_j \zeta(\hat{\mathbf{w}}_{MVDR, I+n}^{(j-1)}; \tilde{\mathbf{r}}_j) \quad (49)$$

converges w.p. 1 to  $\mathbf{w}'_{MVDR, I+n} = \mathbf{R}_{I+n}^{-1} \mathbf{V}_0$ , where  $\zeta(\mathbf{w}; \tilde{\mathbf{r}}_j)$  is defined by (48) and  $\{\alpha_j\}_j$  is a sequence of positive numbers such that  $\sum_{j=1}^{+\infty} \alpha_j = \infty$  and  $\sum_{j=1}^{+\infty} \alpha_j^2 < \infty$ .  $\square$

The initial values of  $\mathbf{v}_0$  and  $\mathbf{w}_{MVDR, I+n}^{(0)}$  are the corresponding values obtained at the end of the training period.

In Fig. 3 we compare the BER performance of the MVDR and AV filter estimators evaluated using  $\hat{\mathbf{R}}_{I+n}$  with the performance of their  $\hat{\mathbf{R}}$ -based counterparts (denoted as traditional). The system setup is the same as in Fig. 2. The AV algorithm is switched to decision directed mode after 50 training samples while the MVDR algorithm after 300 training samples. The performance of the ideal MVDR receiver is included as a reference point. It is evident that the use of  $\hat{\mathbf{R}}_{I+n}$  in evaluating the filter estimators leads to significant BER performance improvements.

## REFERENCES

- [1] M. L. Honig, U. Madhow and S. Verdu, "Blind adaptive multiuser detection," *IEEE Trans. Inform. Theory*, vol. 41, pp. 944-960, July 1995.

- [2] D. A. Pados and S. N. Batalama, "Joint space-time auxiliary-vector filtering for DS/CDMA systems with antenna arrays," *IEEE Trans. Commun.*, vol. 47, pp. 1406-1415, Sept. 1999.
- [3] D. A. Pados and G. N. Karystinos, "An iterative algorithm for the computation of the MVDR filter," *IEEE Trans. Signal Proc.*, submitted.
- [4] A. Kansal, S. N. Batalama, and D. A. Pados, "Adaptive maximum SINR RAKE filtering for DS-CDMA multipath fading channels," *IEEE J. Select. Areas Commun.*, vol. 16, pp. 1765-1773, Dec. 1998.
- [5] T. W. Anderson, *An introduction to multivariate statistical analysis*, J. Wiley and Sons, New York 1958.

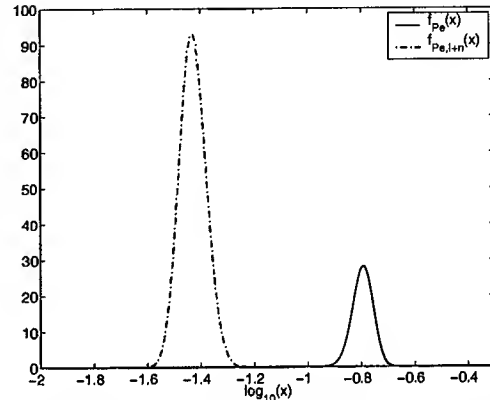


Fig. 1: Density functions of  $P_e(\hat{\mathbf{w}}_{SMI, I+n})$  and  $P_e(\hat{\mathbf{w}}_{SMI})$  for  $N = 200$ .

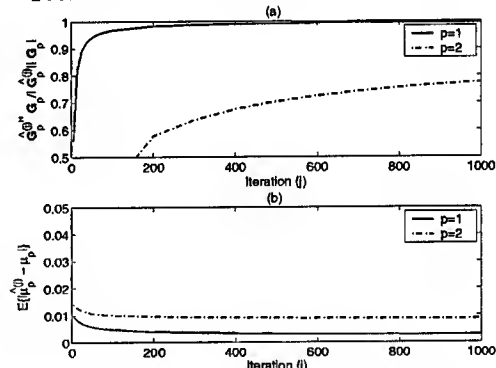


Fig. 2: Convergence of the recursive AV algorithm.

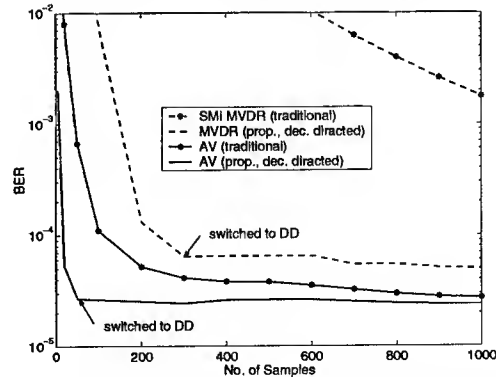


Fig. 3: Bit-error-rate as a function of the data record size  $N$ .

# ON THE EFFECTS OF ROTATING BLADES ON DS/SS COMMUNICATION SYSTEMS

Yimin Zhang<sup>†</sup>, Moeness G. Amin<sup>†</sup>, and Vincent Mancuso<sup>‡</sup>

<sup>†</sup> Department of Electrical and Computer Engineering  
Villanova University, Villanova, PA 19085, USA  
E-mail: {zhang,moeness}@ece.vill.edu

<sup>‡</sup> Boeing Helicopter Division  
Mail Stop P32-21, P. O. Box 16858  
Philadelphia, PA 19142-0858, USA  
E-mail: vincent.mancuso@phl.boeing.com

## ABSTRACT

*Direct-sequence spread-spectrum (DS/SS) techniques are widely used in military and commercial communication systems as well as in the global positioning system (GPS). In this paper, we consider the Doppler effect of rotating blades in a helicopter on the DS/SS communications. We analyze the performance of DS/SS communications under such multipath environment, by taking into consideration both Doppler fading and time delays. The model describing the effects of rotating blades on the desired signal is established, and the system performance is analyzed. It is shown that, for this specific application, the communication channel is not Rayleigh and does not resemble any of the commonly assumed fading models in wireless communications.*

## 1. INTRODUCTION

Direct-sequence spread-spectrum (DS/SS) techniques are widely used in military and commercial communication systems, as well as in the global positioning system (GPS) [1, 2, 3]. In this paper, we focus on airborne antennas for satellite communications mounted on a helicopter. The performance of these systems are affected by multipaths with severe Doppler fading. Although the communication systems may also suffer from the Doppler effects caused by the motion of the helicopter itself, significant signal distortion is induced due to the rotating blades.

The signal received at the airborne antenna is usually a combination of a direct path and local scatterers formed as a result of the signal reflections from the rotating blades. The periodic and time-varying rotational motion of the blades makes the spatial signature of the received scatters time-varying. Rotating blades contribute continuous positive and negative frequency shifts.

The effect of rotating blades on the echo spectrum for purpose of radar target detection was investigated in [4, 5]. In [4], it is pointed out that the signal scattered at the rotating blades yields both amplitude modulation (AM) and frequency modulation (FM). In [5], the radar cross-section spectra of rotating multiple blades is investigated. All these

contributions have considered the problem from the point of view of radar detection of the scattered echo from rotating blades. However, the rotor blade motion impairments of the signal waveforms for wireless communications has not been addressed or investigated.

In this paper, we analyze the performance of DS/SS communications under severe multipath environment, by taking into account the effect of the Doppler fading caused by the rotating blades. Some typical parameters are used to illustrate the effect of time delay and the Doppler frequency shift.

## 2. SIGNAL MODEL

For a single user case, the noise-free signal at the radio frequency (RF) is expressed as

$$x(t) = y(t)e^{j\omega_c t}, \quad (1)$$

where  $\omega_c$  is the carrier radian frequency, and  $y(t)$  is the baseband version of the transmitted signal, which is modeled as

$$y(t) = p(t)d(t), \quad (2)$$

where  $p(t)$  and  $d(t)$  are the spreading waveform and the data-modulated signal waveform, respectively, expressed as

$$p(t) = \sum_{n=-\infty}^{\infty} \sum_{l=1}^{L_c-1} c(n;l)\delta\left(\frac{t}{T_c} - l - nL_c\right) \quad (3)$$

and

$$d(t) = \sum_{n=-\infty}^{\infty} b(n)\delta\left(\frac{t}{T} - n\right). \quad (4)$$

In the above equations,  $b(n)$  is the information symbol,  $c(n;l) \in \{+1, -1\}$  is the aperiodic spreading code at the  $n$ th symbol and the  $l$ th chip,  $L_c$  is the number of chips per symbol,  $T$  and  $T_c$  are the symbol and chip durations, respectively, and

$$\delta(t) = \begin{cases} 1 & 0 \leq t < 1 \\ 0 & \text{elsewhere.} \end{cases}$$

### 3. SCATTERING EFFECTS OF ROTATING BLADES

#### 3.1. Scattering Model

The signal arriving at a point on a blade, illustrated in Fig. 1, is expressed in a general form as

$$x_m(t) = x(t - \tau(\underline{a}_m(t)))A(\underline{a}_m(t))g(\underline{a}_m(t)) \quad (5)$$

where  $\underline{a}_m(t)$  is a vector denoting the position of interest at the  $m$ th blade,  $m = 0, 1, \dots, M-1$ ,  $\tau(\underline{a}_m(t))$  is the time delay,  $A(\underline{a}_m(t))$  is a real scalar representing the scattering loss, and  $g(\underline{a}_m(t))$  is the phase term caused by the Doppler effect. Typically, the number of blades  $M$  is 3 to 5. For notation simplicity, we abbreviate  $\tau(\underline{a}_m(t))$  as  $\tau_m(t)$ , and  $g(\underline{a}_m(t))$  as  $g_m(t)$ .

The time delay  $\tau_m(t)$  is given by  $\Delta l(\underline{a}_m(t))/c$  where  $c$  is the speed of light and  $\Delta l(\underline{a}_m(t)) \triangleq L_1 + L_2 - L_0$ , which represents the additional distance traveled by the multipath relative to the direct path (we use  $L_1$  and  $L_2$  for notation simplicity although both parameters are a function of vector  $\underline{a}_m(t)$ ). The Doppler term in model (5) is expressed as

$$g_m(t) = \exp \left[ -j \frac{2\pi}{\lambda} r \left( \vec{v}(\underline{a}_m(t)) \cdot \vec{l}_1(\underline{a}_m(t)) + \vec{v}(\underline{a}_m(t)) \cdot \vec{l}_2(\underline{a}_m(t)) \right) \right], \quad (6)$$

where  $\lambda$  is the wavelength at the radio frequency, and  $r$  is the distance between the reflection point and the center of the blade, as shown in Fig. 1. This figure illustrates some key parameters in the underlying problem. In equation (6),  $\vec{v}(\underline{a}_m(t))$  is the unit-norm vector describing the direction of the blade movement.  $\vec{l}_1(\underline{a}_m(t))$  and  $\vec{l}_2(\underline{a}_m(t))$  are the unit-norm vectors along the line connecting the point of scattering and the source, and that connecting the point of scattering and the receiving antenna, respectively. Further, " $\cdot$ " denotes the inner product of two vectors. In the scenario considered, the source is located in the far field of the blades, and

$$\vec{v}(\underline{a}_m(t)) \cdot \vec{l}_1(\underline{a}_m(t)) \approx \cos(\alpha) \sin \left( \omega_r t + \frac{2m\pi}{M} \right), \quad (7)$$

where  $\omega_r$  is the rotation radian frequency, and  $\alpha$  is the angle between the line, connecting the source and the center of the rotor, and the plane of the rotor. Both  $\omega_r$  and  $\alpha$  are considered constant over the observation period.

When the receiving antenna is positioned close to the center of the rotor,  $\vec{v}(\underline{a}_m(t))$  and  $\vec{l}_2(\underline{a}_m(t))$  become nearly orthogonal. In this case,  $\vec{v}(\underline{a}_m(t)) \cdot \vec{l}_2(\underline{a}_m(t))$  is negligible, and the Doppler effect in equation (6) can be simplified to

$$g_m(t) \approx \exp \left[ -j \frac{2\pi}{\lambda} r \cos(\alpha) \sin \left( \omega_r t + \frac{2m\pi}{M} \right) \right] \triangleq \exp[j\phi_m(t)]. \quad (8)$$

To further simplify the analysis, we make the following assumptions regarding the rotor blades.

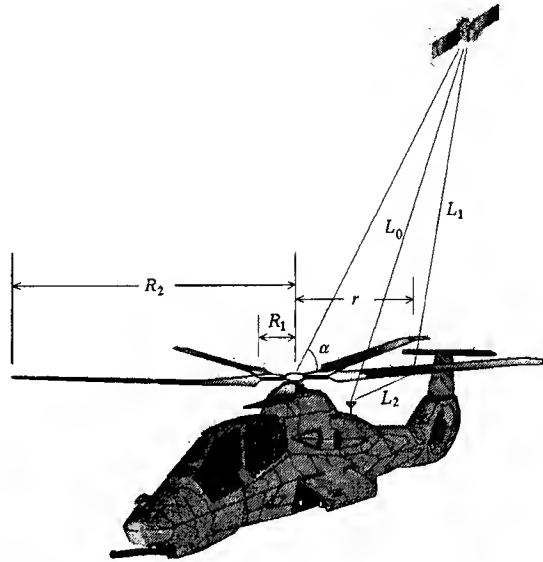


Fig. 1 Parameters of the blades.

- A1) Each blade acts as a homogeneous, linear, rigid antenna.
- A2) Each blade is always visible to both the source and the airborne antenna, i.e., there is no shielding of the blades.
- A3) The near field effect and the secondary scattering effect are not considered.

The overall received signal is an integral of equation (5) over the extent of the blade and is given by (the scattering loss density  $A$  is used instead of  $A(t, r)$  because of assumption A1)

$$\begin{aligned} x_r(t) &= x(t) + \sum_{m=0}^{M-1} \int_{R_1}^{R_2} x_m(t) dr \\ &= x(t) + \sum_{m=0}^{M-1} \int_{R_1}^{R_2} x(t - \tau_m(t)) A e^{j\phi_m(t)} dr, \end{aligned} \quad (9)$$

where  $R_1$  and  $R_2$  are, respectively, the distance of the blade roots and that of the blade tips, from the center of the rotation. The first term at the right hand of equation (9) stands for the contribution of the direct path, whereas the second term is the contribution of the scatters.

The baseband signal associated with equation (9), tak-

ing into account the effect of noise term,  $n(t)$ , is given by

$$\begin{aligned}
y_r(t) &= y(t) \\
&+ \sum_{m=0}^{M-1} \int_{R_1}^{R_2} y(t - \tau_m(t)) A e^{j[\phi_m(t) - \omega_c \tau_m(t)]} dr \\
&+ n(t) \\
&= p(t)d(t) \\
&+ \sum_{m=0}^{M-1} \int_{R_1}^{R_2} p(t - \tau_m(t)) d(t - \tau_m(t)) A e^{j\tilde{\phi}_m(t)} dr \\
&+ n(t),
\end{aligned} \tag{10}$$

where

$$\begin{aligned}
\tilde{\phi}_m(t) &= \phi_m(t) - \omega_c \tau_m(t) \\
&= \phi_m(t) - 2\pi \Delta l(\underline{a}_m(t))/\lambda \\
&= \phi_m(t) - 2\pi(L_1 + L_2 - L_0)/\lambda
\end{aligned} \tag{11}$$

is the combined phase term caused by the Doppler effect and the propagation delay.

Despreading the signal over the  $n$ th symbol yields

$$\begin{aligned}
z_r(n) &= \frac{1}{T} \int_{nT}^{(n+1)T} y_r(t) p(t) dt \\
&= \frac{1}{T} \int_{nT}^{(n+1)T} d(t) dt \\
&+ \frac{1}{T} \sum_{m=0}^{M-1} \int_{nT}^{(n+1)T} \int_{R_1}^{R_2} p(t - \tau_m(t)) p(t) \\
&\quad \times d(t - \tau_m(t)) A e^{j\tilde{\phi}_m(t)} dr dt \\
&+ \frac{1}{T} \int_{nT}^{(n+1)T} n(t) p(t) dt \\
&= b(n) \\
&+ \frac{1}{T} \sum_{m=0}^{M-1} \int_{nT}^{(n+1)T} \int_{R_1}^{R_2} p(t - \tau_m(t)) p(t) \\
&\quad \times d(t - \tau_m(t)) A e^{j\tilde{\phi}_m(t)} dr dt \\
&+ n_z(t),
\end{aligned} \tag{12}$$

where  $n_z(t)$  is the noise component after despreading. The first term at the right hand of equation (12) is the received signal from the direct path, whereas the second term is the contribution of the scattered signals, upon despreading.

### 3.2. Discussions on the Scattering Components

The significance of the effect of rotating blades highly depends on various parameters, such as the dimension of the blades, the position of the antenna, the RF frequency, the angle  $\alpha$ , and the symbol and chip rates. Below we use some typical parameters [2, 6], given in Table 1, to illustrate interesting model properties. Different values of  $\alpha$  are considered, and their effects on the Doppler shift and the bit

error rate (BER) are demonstrated. The role of time delay, Doppler shift, and the equivalent channel characteristics is examined. It is important to note that the arguments presented below are based on the specific values listed in Table 1, and may change with signal coding/modulation and rotorcraft structure and dimension.

#### Parameter Variation

When considering a symbol period, the instantaneous Doppler frequency shift for a point can be assumed unchanged, since  $\omega_r T$  is often very small. For example, when the symbol rate is 100 kbauds,  $\omega_r T = 8\pi \times 10^{-5} = 2.51 \times 10^{-4} \text{ (rad)} = 0.0144^\circ$ .

At the blade tips, the distance traveled over this period is  $\omega_r T R_2 = 2.51 \times 10^{-4} \times 7.5 \approx 1.8 \times 10^{-3} \text{ (m)}$ , which is relatively small compared with the wavelength (0.03 m at 10 GHz RF). Therefore, the position of the blades can be considered unchanged over a symbol period. However, this small difference in position may result in propagation phase change.

Table 1: Typical Parameters Considered

Parameter	Notation	Typical	value
Radio frequency	$\omega_c/2\pi$	10	GHz
Chip rate	$1/T_c$	10	Mcps
Symbol rate	$1/T$	100	kbauds
Diameter of blades	$2R_2$	15	m
Rotation speed	$\omega_r/2\pi$	4	r/s

#### Time Delay Consideration

We assume that the antenna is located close to the center of the rotor. In this case, the maximum possible delay is  $\Delta l_{max} = (1 + |\cos(\alpha)|) R_2$ . Consider a typical scenario where  $R_2$  is 7.5 meters. Then, the corresponding maximum possible time delay  $\tau_{max} = \Delta l_{max}/c$  is 33.6 ns in the case when  $\alpha = 70^\circ$ .

If the chip rate is 10 Mcps, the chip period is 100 ns, which is about three times the maximum possible time delay. Therefore, the time delay cannot be totally ignored, but its effect may not be significant.

The relative time delay with respect to the chip period becomes larger as the chip rate increases. However, as it is clear from equation (12), in the case when the maximum time delay is larger than the chip period, the multipath components whose delays exceed the chip period will be discriminated at the receiver by the virtue of despreading. Therefore, the maximum time delay to be considered is the chip period.

#### Doppler Effect Consideration

One of the important parameters in the underlying rotor scattering problem is the maximum Doppler frequency. The instantaneous Doppler frequency shift of the scattered signals is the derivative of the phase defined in equation (8), and



is given by

$$\begin{aligned}\Delta f_m(t) &= \frac{d\phi_m(t)}{2\pi dt} \\ &= -\frac{r\omega_r}{\lambda} \cos(\alpha) \cos\left(\omega_r t + \frac{2m\pi}{M}\right).\end{aligned}\quad (13)$$

Accordingly, the maximum instantaneous frequency shift of the scattered signals is

$$\Delta f_{m,max}(\alpha) = \max_{t,m,r} \Delta f_m(t) = \frac{R_2\omega_r}{\lambda} \cos(\alpha). \quad (14)$$

For example, consider the case in which the blades rotate at 240 rpm or 4 r/s,  $\omega_r = 8\pi$ . At the RF frequency 10 GHz,  $\lambda = 0.03$  m, and the upper bound of the Doppler frequency is obtained at  $\alpha = 0$  and equal to  $\Delta f_{m,max}(0) = 7.5 \times 8\pi/0.03 = 6.28$  kHz. Due to the  $\cos(\alpha)$  term, the maximum Doppler frequency shift  $\Delta f_{m,max}(\alpha)$  is often smaller than the above bound.

Although the maximum Doppler frequency is much smaller than the chip rate, it is at a comparable order to the symbol rate. It is clear from equation (12) that, since an integration is performed over a symbol period, the contribution of the scattering multipaths with Doppler frequencies higher than the symbol rate is small.

Fig. 2 shows the instantaneous Doppler frequency shift  $\Delta f_m(t)$  at the blade tips ( $r = R_2 = 7.5$  m) for the 0th blade ( $m=0$ ) versus time in terms of the number of symbols. The results are shown for a period of one rotation cycle (0.25 sec =  $25,000T$ ), where the angle  $\alpha$  takes the values  $0^\circ$ ,  $45^\circ$ ,  $70^\circ$ , and  $90^\circ$ . The Doppler frequency shift is proportional to  $\omega_c$ ,  $\omega_r$ , and  $r$ . As such, any change in these parameters leads to a linear change of the Doppler frequency shift.

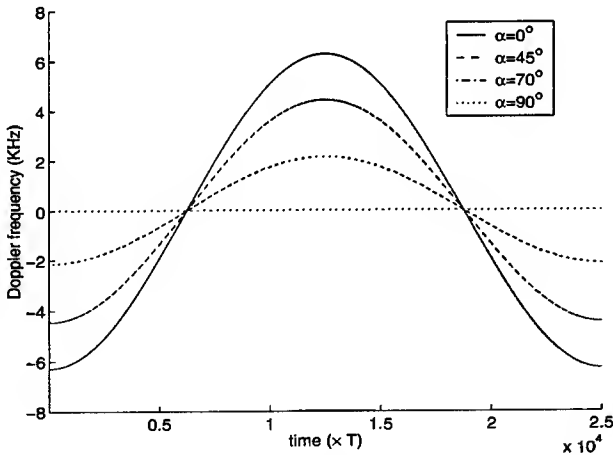


Fig. 2 Doppler frequency shift vs. time ( $m=0$ ,  $r=7.5$  m).

#### Channel Characteristics

To examine the effects of the scattering in equation (12), we note the fact that  $\tilde{\phi}_m(t)$  can be considered constant over

a symbol period. Define

$$\rho(\tau_m(t)) = \frac{1}{T} \int_{nT}^{(n+1)T} p(t - \tau_m(t))p(t)dt \quad (15)$$

as the correlation function of the chip waveform. Then, the scattering component is approximated by

$$\begin{aligned}z_r^{(s)}(n) &\triangleq \frac{1}{T} \sum_{m=0}^{M-1} \int_{nT}^{(n+1)T} \int_{R_1}^{R_2} p(t - \tau_m(t))p(t) \\ &\quad \times d(t - \tau_m(t)) A e^{j\tilde{\phi}_m(t)} dr dt \\ &\approx Ab(n) \sum_{m=0}^{M-1} \int_{R_1}^{R_2} \rho(\tau_m(t)) e^{j\tilde{\phi}_m(t)} dr \\ &\triangleq \xi(t)b(n),\end{aligned}\quad (16)$$

where

$$\xi(t) = A \sum_{m=0}^{M-1} \int_{R_1}^{R_2} \rho(\tau_m(t)) e^{j\tilde{\phi}_m(t)} dr \quad (17)$$

is the channel response of the scattering component contribution.

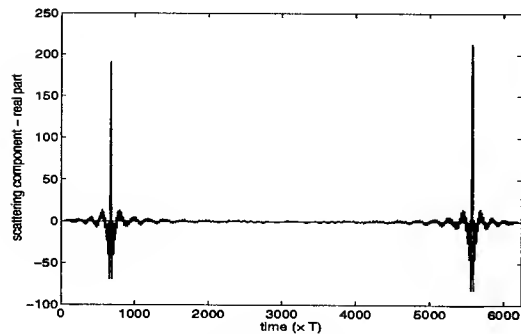
Examples of  $\xi(t)$  (real part) are shown in Fig. 3 for different values of  $\alpha$ , where we have assumed that the antenna is located close to the center of the rotors,  $M = 4$ ,  $\rho(\tau_m(t)) = 1 - |\tau_m(t)|/T_c$ , and  $R_1 = 1$  m. The results are shown for 6250 symbols, or equivalently over one full period of  $\xi(t)$ , given by  $1/(Mf_r)$ . We set  $A = 1/\lambda$  for normalization. It is evident that, when  $\alpha$  is small,  $\xi(t)$  demonstrates large peak values. On the other hand, when  $\alpha$  is large (close to  $90^\circ$ ), the result becomes very small. The reason is discussed below.

Naturally, when  $\alpha$  is close to  $90^\circ$ , the effect of Doppler frequency shift is small. Since  $L_0 \approx L_1$ , then  $L_2$  becomes the dominant parameter in defining the time delay, and subsequently the propagation phase. This same property causes the phase to be periodic over  $r$ . Because of the periodicity, averaging over  $r$  will then lead to small values. On the other hand, when  $\alpha$  is small, the two contributions to the phase by the Doppler frequency shift and the propagation delay from  $L_0 - L_1$  become comparable. Large peaks appear when these components are resonant with respect to  $r$ .

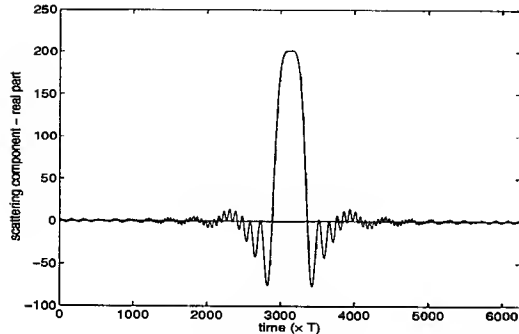
#### 4. BER PERFORMANCE

Computer simulations are performed to illustrate the scattering effect on the BER performance. Both the symbol and chip modulations are assumed to be binary phase shift keying (BPSK). The typical parameters listed in Table 1 are used. The noise is assumed to be a white Gaussian random process.

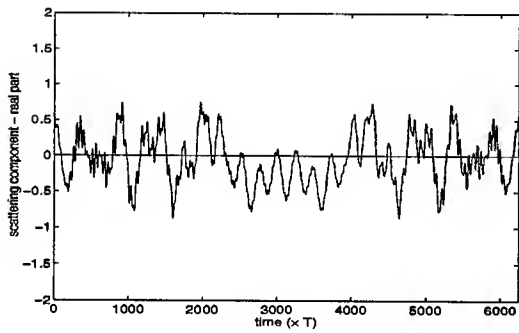
Fig. 4 shows the BER performance versus the input signal-to-noise ratio (SNR) for different values of  $A$  and  $\alpha$ . When  $A$  is large ( $A \geq 10^{-2}/\lambda$  in this figure) and  $\alpha$  is small ( $\alpha \leq 45^\circ$  in this figure), the BER takes a high value which slowly decrease with increased input SNR. This is the impact of large scattering from the rotating blades. On the other hand, when  $A$  is small ( $A = 10^{-3}/\lambda$  in this figure) or  $\alpha$  is close to  $90^\circ$  ( $\alpha = 70^\circ$  in this figure), the effect of the rotating blades is negligible.



(a)  $\alpha = 30^\circ$



(b)  $\alpha = 45^\circ$



(c)  $\alpha = 70^\circ$

Fig. 3 Scattering component  $\xi(t)$  vs. time.

## 5. CONCLUSION

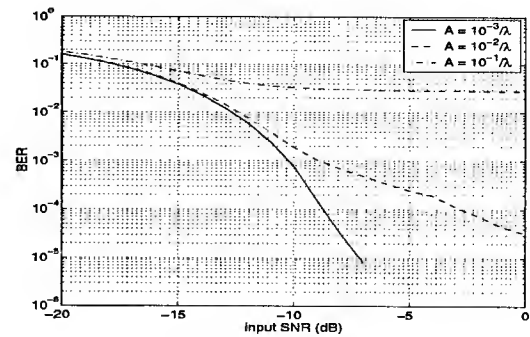
We have analyzed the channel characteristics and the performance of direct-sequence spread-spectrum (DS/SS) communications under the multipath environment caused by the rotating blades in a helicopter, by taking both effect of the Doppler fading and the time delays.

## ACKNOWLEDGMENT

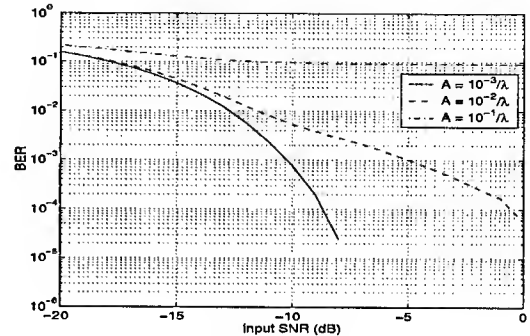
The authors would like to thank Weifeng Mu for his contribution to the scattering model.

## REFERENCES

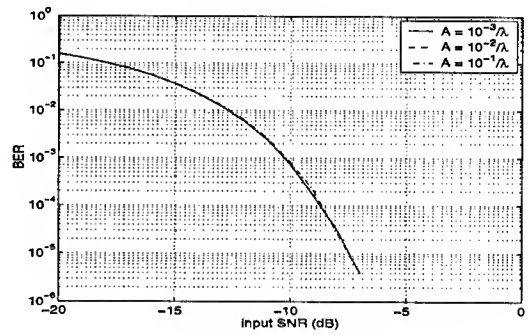
- [1] R. C. Dixon, *Spread Spectrum Systems*, John Wiley, 1984.



(a)  $\alpha = 30^\circ$



(b)  $\alpha = 45^\circ$



(c)  $\alpha = 70^\circ$

Fig. 4 BER vs. input SNR.

- [2] R. C. Dixon, *Spread Spectrum Systems with Commercial Applications*, 3rd Ed., John Wiley, 1994.
- [3] B. W. Parkinson, ed, *Global Positioning System: Theory and Applications*, vol. I, AIAA, 1996.
- [4] J. Martin and B. Mulgrew, "Analysis of the theoretical radar return signal from aircraft propeller blades," in *Proc. 1990 IEEE. Int. Radar Conf.*, Washington D. C., May 1990.
- [5] S.-S. Bor, T.-L. Yang, and S.-Y. Yang, "Radar cross-sectional spectra of rotating multiple skew-plated metal fan blades by physical optics/physical theory of diffraction, equivalent currents approximations," *Jpn. J. Appl. Phys.*, vol. 31, no. 5A, pp. 1549-1554, May 1992.
- [6] W. Z. Stepniewski and C. N. Keys, *Rotary-Wing Aerodynamics*, Dover, 1984.

# JOINT SYNCHRONIZATION AND SYMBOL DETECTION IN ASYNCHRONOUS DS-CDMA SYSTEMS

Francesc Rey, Gregori Vázquez, Jaume Riba

Department of Signal Theory and Communications, Polytechnic University of Catalonia  
UPC Campus Nord - Mòdul D5, c/Jordi Girona 1-3, 08034 Barcelona (Spain)  
e-mail:{frey,gregori,jriba}@gps.tsc.upc.es

## ABSTRACT

Design accurate estimators which also consider the noise term in low  $SNR$  scenarios is paramount to achieve optimal solutions and to obtain precise symbol detectors. Particularly, this paper estimates the propagation delays focusing on asynchronous  $DS-CDMA$  systems. The proposed *Minimum Conditioned Variance (MCV)* is the choice in noisy environments, implementing the best linear detector of the transmitted symbols under a minimum mean-square error criterion. The result is an estimator that improves the *conditional ML (CML)* solution when noise is not negligible, and attains the derived *Gaussian Unconditional Cramér-Rao Bound (UCRB)* in the whole  $E_b/N_0$  range as classical *Gaussian Unconditional ML (UML)* does. Consequently, the proposed *MCV* estimator, becomes an optimal quadratic solution achieving similar features than *UML* in a straightforward way, and with no assumptions on the signal statistics.

## 1. INTRODUCTION

In digital communications, the knowledge of certain parameters as for example the phase and carrier frequency or the propagation delay, are paramount to get a reliable detection of the transmitted symbols. Focusing on multi-user  $DS-CDMA$  systems, an accurate estimation of the propagation delays for all users is essential. Otherwise, the performance of the multi-user detector is rapidly decreased by means of multiple-access interference (*MAI*), as has been widely studied in the literature [1], [2]. Accordingly, this paper addresses a multi-parametric estimator intended for the multi-user synchronization and symbol detection, with high performance in low  $SNR$  scenarios. Nevertheless, the proposed algorithm is not restricted to multi-user synchronizers, and can be also extended to other estimation problems, like frequency synchronization in *OFDM* and *Multi-Carrier* schemes.

Maximum Likelihood (*ML*) formulation has been usually employed to design timing estimators. Classically, *Unconditional ML (UML)* algorithms have been developed in

the field of digital communications modeling the transmitted symbols as stochastic processes. Nevertheless, in order to obtain feasible mathematical expressions, *UML* estimators make some assumptions on the gaussianity of signal statistics, which is known to be a non-realistic assumption in digital communications, or assumptions on low  $SNR$ , which leads to *self-noise* appreciable when the noise term is negligible. Consequently, the restrictions on *UML* motivated the introduction of *deterministic* or *conditional ML (CML)*, which considers the transmitted symbols as deterministic unknown parameters. This formulation has been applied by Stoica and Nehorai [3] in sensor array processing to perform *DOA* estimation, and more recently the same principle has been applied to frequency and timing estimation [4]-[7]. The *CML* solution does not present *self-noise*, is robust in *near-far* scenarios, and provides a high performance at high  $SNR$ 's. Nevertheless it is not an optimal solution in noisy scenarios with low  $SNR$ .

The proposed *Minimum Conditioned Variance (MCV)* method, addressed in this paper, mitigates the *CML* estimation drawbacks at low  $SNR$  scenarios considering the impact of the noise, and becomes the deterministic solution at high  $SNR$ . Although the derived *MCV* becomes biased, the bias value can be estimated and next subtracted to obtain an unbiased estimator. The result is an estimator that attains the lower *Gaussian Unconditional Cramér-Rao Bound UCRB* in the whole  $E_b/N_0$  range, as *Gaussian UML* does. Accordingly, *MCV* becomes an optimal quadratic estimator with no assumptions on the signal statistics.

This paper is organized as follows. Next section describes the discrete-time signal model, and obtains a structured matrix expression containing the parameters to estimate. Section 3 describes the *CML* formulation and justifies under which conditions the deterministic criterion does not become feasible. Afterwards, section 4 introduces the *Minimum Conditioned Variance* method as choice, and derives its gradient expression. Furthermore, a detailed study of the proposed estimator shows it is biased and consequently a modified unbiased estimator is proposed. Next, section 5 derives the *UCRB* which is used as a benchmark, at high and low  $SNR$ 's, to the performance of the proposed multi-user delay estimator. Finally last section presents some simulation results proving the proposed *MCV* outperforms *CML*, attaining the *UCRB* and reducing the Bit-Error Rate *BER* in symbol detection.

This work has been supported by: TIC98-0412, TIC98-0703, TIC99-0849 (CICYT) and CIRIT/Generalitat de Catalunya 1998SGR-00081.

## 2. DISCRETE-TIME SIGNAL MODEL

The described model considers a  $K$  user asynchronous *DS-SS* system operating in a multipath environment. The received signal contains the superposition of  $K$  active users:

$$r(t) = \sum_{k=1}^K s^k(t - \tau_k) + w(t) \quad (1)$$

where  $s^k(t)$  denotes the  $k$ -user received baseband signal,  $\tau_k$  its the propagation delay, and  $w(t)$  represents the received AWGN noise term with zero mean and variance  $\sigma_w^2$ . For each user the received baseband signal is modeled as:

$$s^k(t) = \sum_{n=-\infty}^{\infty} d_n^k e^{j\theta_k} g^k(t - nT) \quad (2)$$

where  $g^k(t)$  represents the  $k$ -user received signature,  $T$  is the bit duration,  $d_n^k$  are the transmitted information bits, and  $\theta_k$  the received carrier phase. Moreover, considering the presence of a propagation multipath channel with baseband impulse response  $h^k(t)$ , the  $k$ -user received signature is given by a distorted version of the transmitted spreading waveform  $c^k(t)$  as:

$$g^k(t) = c^k(t) * h^k(t) \quad (3)$$

Finally the received signal as a function of the user's signatures is given by:

$$r(t) = \sum_{k=1}^K \sum_{n=-\infty}^{\infty} d_n^k e^{j\theta_k} g^k(t - nT - \tau_k) + w(t) \quad (4)$$

The algorithm is derived in a discrete-time signal model by sampling the received waveform at  $N_{sc}$  samples per chip. Choosing the sampling frequency as  $f_s = 1/T_s$ , where  $T_s$  is the sampling period, and collecting  $2M + 1$  samples of  $r(nT_s)$ , the vector  $\mathbf{r}$  can be defined as:

$$\mathbf{r} = [r(-MT_s) \quad \dots \quad r(0) \quad \dots \quad r(MT_s)]^T \quad (5)$$

At this point equation (4) can be expressed following the matrix signal model:

$$\mathbf{r} = \mathbf{A}(\tau)\mathbf{x} + \mathbf{w}^* \quad (6)$$

The set of unknown parameters (i.e. the transmitted symbols and phase errors) for  $k$ -user define the vector  $\mathbf{x}^k$ :

$$\mathbf{x}^k = [d_{-L}^k e^{j\theta_k} \quad \dots \quad d_0^k e^{j\theta_k} \quad \dots \quad d_L^k e^{j\theta_k}]^T \quad (7)$$

where the number of transmitted symbols  $N_s = 2L + 1$ . Finally, stacking all users, the nuisance parameter vector  $\mathbf{x}$  is defined as follows:

$$\mathbf{x} = [\mathbf{x}^{1T} \quad \mathbf{x}^{2T} \quad \dots \quad \mathbf{x}^{KT}]^T \quad (8)$$

<sup>\*</sup>The channel coefficients are assumed to be known or previously estimated (e.g. [8])

On the other hand the model transfer matrix, denoted as  $\mathbf{A}(\tau)^{\dagger}$ , contains the user signatures, and the parameters to estimate  $\tau_k$ :

$$\mathbf{A} = \mathbf{A}(\tau) = [\mathbf{A}^1(\tau_1) \quad \mathbf{A}^2(\tau_2) \quad \dots \quad \mathbf{A}^K(\tau_K)] \quad (9)$$

$$\begin{aligned} \mathbf{A}^k(\tau_k) &= [\mathbf{a}_0^k \quad \mathbf{a}_1^k \quad \dots \quad \mathbf{a}_{N_s-1}^k] \\ \mathbf{a}_n^k &= [g^k(-MT_s - nT - \tau_k) \quad \dots \\ &\quad g^k(MT_s - nT - \tau_k)]^T \end{aligned}$$

where the columns of  $\mathbf{A}^k(\tau_k)$  are scrolled versions of the  $k$ -user signature delayed  $\tau_k$ .

A more detailed model of matrix  $\mathbf{A}^k(\tau_k)$  will be constituted by the product of two matrices:

$$\mathbf{A}^k(\tau_k) = \mathbf{H}^k(\mathbf{h}_k) \mathbf{C}^k(\tau_k) \quad (10)$$

Matrix  $\mathbf{H}^k(\mathbf{h}_k)$  is a Sylvester or convoluting matrix modeling the channel distortion, whose columns are the  $k$ -user impulsive channel response coefficients. On the other hand, matrix  $\mathbf{C}^k(\tau_k)$  will be obtained by the  $k$ -user spreading code delayed  $\tau_k$ .

## 3. THE CML FORMULATION

The cost function in *CML* estimation for the signal model in (6) is derived from the joint *ML* cost function that is formulated as:

$$\Lambda(\mathbf{r}/\tau, \mathbf{x}) = \frac{1}{(\pi\sigma_w^2)^M} e^{-\frac{1}{\sigma_w^2} \|\mathbf{r} - \mathbf{A}\mathbf{x}\|^2} \quad (11)$$

The *ML* function depends on the parameter estimation vector  $\tau$  and also on the vector  $\mathbf{x}$ . Notice that vector  $\mathbf{x}$  contains the set of unknown parameters and thus it is necessary to take some considerations on this vector. The joint  $\tau, \mathbf{x}$  estimation could be the solution, but it is discarded because it is computationally complex, and alternative algorithms only focusing on the  $\tau$  vector estimation are proposed. Classically, *UML* solution computes the expectation of the joint *ML* function with respect to the nuisance parameters:

$$\Lambda_{UML}(\mathbf{r}/\tau) = E_{\mathbf{x}} \{\Lambda(\mathbf{r}/\tau, \mathbf{x})\} \quad (12)$$

In general the expectation  $E_{\mathbf{x}}$  in (12) is quite difficult to obtain, and in practice only an approximation of the likelihood function in low *SNR* scenarios is approached.

Previous limitations motivate the use of the *CML* solution. This method considers the nuisance parameters as deterministic, and thus they can be substituted by its estimation keeping fixed  $\tau$  vector. The *ML* estimation of  $\mathbf{x}$ , when no restrictions are imposed on it, can be obtained as:

$$\hat{\mathbf{x}}_{ML} = \mathbf{A}^{\#} \mathbf{r} \quad (13)$$

where  $\mathbf{A}^{\#}$  is the Moore-Penrose pseudo-inverse. Once the nuisance vector  $\mathbf{x}$  is estimated, the compressed *ML* function to maximize, which only depends on the parameter vector  $\tau$ , is obtained by replacing (13) in (11). And finally the

<sup>†</sup>Hereafter the dependence on vector  $\tau$  will be suppressed for simplicity

derived log-likelihood function to minimize (omitting irrelevant constants) is given by:

$$\min_{\tau} L_{CML}(\mathbf{r}/\tau) = \text{tr} \{ \mathbf{P}_A^\perp \hat{\mathbf{R}} \} \quad (14)$$

where  $\mathbf{P}_A^\perp = \mathbf{I} - \mathbf{A}\mathbf{A}^\#$  is the projection matrix onto the orthogonal subspace defined by  $\mathbf{A}$ , and  $\hat{\mathbf{R}} = \mathbf{r}\mathbf{r}^H$ .

To minimize (14) a gradient algorithm may be used. The gradient in conditional ML was derived by Viberg, Ottersten and Kailath [9] in the context of array processing for DOA estimation. In our delay estimation problem this gradient can be expressed as:

$$g_{c_i} = -2\text{Re} \{ (\mathbf{r}^H \mathbf{P}_A^\perp \mathbf{D}_i) (\mathbf{A}^\# \mathbf{r}) \} \quad (15)$$

where  $\mathbf{D}_i = \frac{\partial}{\partial \tau_i} \mathbf{A}$ .

A more accurate study of the gradient expression shows that it is computed by the product of two terms. The first term is  $(\mathbf{r}^H \mathbf{P}_A^\perp \mathbf{D}_i)$  and justifies the proposed algorithm to be *self-noise* free. Considering a noiseless environment, and the absence of delay errors, vector  $\mathbf{r}$  will be contained in the signal subspace generated by the  $\mathbf{A}$  matrix columns. Thus, the projection matrix  $\mathbf{P}_A^\perp$ , which does not appear in the classical unconditional approach, acts as a zero-forcer placed at the output of the derivative matched filter  $\mathbf{D}_i$ . As a result, the estimator ensures in all cases a *self-noise* free solution:  $(\mathbf{r}^H \mathbf{P}_A^\perp \mathbf{D}_i) = 0$ .

The second term  $(\mathbf{A}^\# \mathbf{r})$  corresponds to the ML estimation of the unconstrained vector  $\mathbf{x}$ . Notice that this expression is the decorrelating detector solution, so the algorithm not only estimates the propagation delay but also implements this sub-optimum detector. The presence of this term justifies the proposed solution to be a robust *near-far* estimator. Analyzing the signal model (6) it is observed that the received powers can be introduced in the nuisance parameter vector  $\mathbf{x}$ . Hence, following (13) it is guaranteed that the algorithm will estimate the received power values, justifying the estimator to be insensitive to different power levels.

Nevertheless, the decorrelating detector evidences some difficulties in noisy scenarios. The pseudoinverse, as the ideal zero-forcing solution ZF in equalization, does not take into account the noise term. Accordingly, when the transfer matrix  $\mathbf{A}$  eigenvalue spreading, defined as:

$$\chi = \frac{\lambda_{A_{max}}}{\lambda_{A_{min}}} \quad (16)$$

is large enough, the noise term will be extremely increased, becoming the CML method an unacceptable solution in low SNR scenarios, which are common in wideband DS-CDMA systems.

#### 4. MINIMUM CONDITIONED VARIANCE APPROACH

A novel approach is proposed in this paper considering the impact of the noise in the likelihood function, achieving in consequence a more robust estimator in low SNR scenarios. The *Minimum Conditioned Variance* approach (MCV) makes the nuisance parameter estimation as the best linear

estimation under a minimum variance criterion given an observation vector  $\mathbf{r}$ . This estimation is:

$$\begin{aligned} \hat{\mathbf{x}} &= E[\mathbf{x}/\mathbf{r}] = \mathbf{\Gamma} \mathbf{A}^H (\mathbf{A} \mathbf{\Gamma} \mathbf{A}^H + \sigma_w^2 \mathbf{I})^{-1} \mathbf{r} = \mathbf{C} \mathbf{r} \\ \mathbf{C} &= \mathbf{\Gamma} \mathbf{A}^H (\mathbf{A} \mathbf{\Gamma} \mathbf{A}^H + \sigma_w^2 \mathbf{I})^{-1} \\ \mathbf{\Gamma} &= E\{\mathbf{x}\mathbf{x}^H\} \end{aligned} \quad (17)$$

Previous expression belongs to the best linear and non-linear estimator under Gaussian conditions, and only the best linear estimator under non-Gaussian conditions. The new cost function is derived by substituting (17) in equation (11) and it is given by:

$$\min_{\tau} L_{MCV}(\mathbf{r}/\tau) = \|\mathbf{r} - \mathbf{A}\mathbf{C}\mathbf{r}\|^2 \quad (18)$$

At high SNR scenarios  $\mathbf{C}(\sigma_w^2 \rightarrow 0) = \mathbf{A}^\#$  is the pseudo-inverse of  $\mathbf{A}$ , becoming the CML solution. On the other hand, when the contribution of  $\mathbf{A} \mathbf{\Gamma} \mathbf{A}^H$  is negligible in front of  $\sigma_w^2 \mathbf{I}$ ,  $\mathbf{C}$  approaches a bank of matched filters containing all the user signatures:  $\mathbf{C}(\sigma_w^2 \rightarrow \infty) = \sigma_w^{-2} \mathbf{\Gamma} \mathbf{A}^H$ . This second limit is achieved at low SNR when the noise power is much greater than the received signal power for all users. Notice however that, in high *near-far* scenarios, the elements in  $\mathbf{\Gamma}$  associated to the most powerful users will be higher than the noise term. Consequently, in scenarios with low SNR and small *near-far*, the MCV will improve the classical CML solution, whereas in high *near-far* scenarios, MCV will remain close to CML.

To minimize (18) we will follow once again a gradient scheme. The gradient expression in MCV is given by:

$$g_{mcv_i} = -2\text{Re} \left\{ \mathbf{r}^H (\mathbf{I} - \mathbf{A}\mathbf{C})^H \left( \mathbf{D}_i \mathbf{C} + \mathbf{A} \frac{\partial}{\partial \tau_i} \mathbf{C} \right) \mathbf{r} \right\} \quad (19)$$

It results interesting to analyze the behaviour at high and low SNR scenarios. At high SNR  $\mathbf{C} \rightarrow \mathbf{A}^\#$ , and making use of  $\mathbf{P}_A^\perp \mathbf{A} = 0$ , the second term in the previous gradient is asymptotically equal to zero:  $\mathbf{r}^H (\mathbf{I} - \mathbf{A}\mathbf{C})^H \mathbf{A} \frac{\partial}{\partial \tau_i} \mathbf{C} \mathbf{r} = 0$ . Thus the gradient becomes:

$$g_{mcv_i}(\sigma_w^2 \rightarrow 0) \simeq -2\text{Re} \{ \mathbf{r}^H (\mathbf{I} - \mathbf{A}\mathbf{C})^H (\mathbf{D}_i \mathbf{C}) \mathbf{r} \} \quad (20)$$

Likewise, at low SNR's  $\mathbf{C} \rightarrow \sigma_w^{-2} \mathbf{\Gamma} \mathbf{A}^H$ , and the two components in the gradient (19) supply the same value. Hence, the asymptotic gradient derived in noisy environments corresponds with:

$$g_{mcv_i}(\sigma_w^2 \rightarrow \infty) \simeq -\frac{4}{\sigma_w^2} \text{Re} \{ \mathbf{r}^H \mathbf{D}_i \mathbf{\Gamma} \mathbf{A}^H \mathbf{r} \} \quad (21)$$

Notice that the second term can be dropped in both cases without losing information by the gradient.

Finally, for the special case when there is only one parameter to estimate, e.g. timing or frequency estimation in linear and non-linear modulations, another argument to eliminate the second term is detailed in [7] and next outlined. Considering that vector  $\mathbf{r}$  follows a Gaussian distribution, which is known to be a non-realistic assumption in digital communications, the Gaussian UML cost function becomes <sup>†</sup>:

$$\begin{aligned} L_{UMLG}(\mathbf{r}/\tau) &= \mathbf{r}^H \mathbf{R}^{-1} \mathbf{r} \\ \mathbf{R} &= (\mathbf{A} \mathbf{\Gamma} \mathbf{A}^H + \sigma_w^2 \mathbf{I}) \end{aligned} \quad (22)$$

<sup>†</sup>only applicable if  $\mathbf{A}^H \mathbf{A}$  does not depend on the parameter to estimate

and the Gaussian *UML* gradient in previous equation is given by:

$$g_{uml_i} = -2\text{Re} \{ \mathbf{r}^H (\mathbf{I} - \mathbf{A}\mathbf{C})^H \mathbf{D}_i \mathbf{C} \mathbf{r} \} \quad (23)$$

Comparing last equation with (19), a further justification for removing the second term is obtained. Accordingly, in uni-parametric estimators, and assuming a Gaussian distribution for the transmitted symbols, the *MCV* gradient becomes the Gaussian *UML* gradient. Nevertheless, in multi-parametric estimators, the Gaussian *UML* cost function becomes more complex:

$$L_{UMLG}(\mathbf{r}/\tau) = \ln |\mathbf{R}| + \mathbf{r}^H \mathbf{R}^{-1} \mathbf{r} \quad (24)$$

Notice a new term  $\ln |\mathbf{R}|$ , which becomes constant in the uniparametric estimators when  $\mathbf{A}^H \mathbf{A}$  does not depend on the parameter to estimate, is introduced. The gradient expression, derived in [3] cannot be identified with (19) anymore.

After the previous analysis, the *MCV* gradient can be asymptotically rewritten as:

$$g_{mcv_i} \approx -2\text{Re} \{ \mathbf{r}^H (\mathbf{I} - \mathbf{A}\mathbf{C})^H \mathbf{D}_i \mathbf{C} \mathbf{r} \} \quad (25)$$

A more accurate analysis of the previous gradient shows it is biased. It can be seen that in the absence of timing errors the gradient does not become the null vector. Therefore, the bias expression can be obtained computing the gradient expected value when the estimated timing vector equals the real timing vector:

$$\begin{aligned} \text{Bias}_i &= E \{ g_{mcv_i} \} |_{\hat{\tau}=\tau} \\ &= -2\text{Re} \{ \text{Tr} \{ \mathbf{\Gamma} \mathbf{A}_\tau^H (\mathbf{I} - \mathbf{A}_\tau \mathbf{C}_\tau)^H \mathbf{D}_{i_\tau} \} \} \end{aligned} \quad (26)$$

denoting  $\mathbf{A}_\tau$ ,  $\mathbf{C}_\tau$ ,  $\mathbf{D}_{i_\tau}$ , the dependence of matrices on  $\tau$ . Unfortunately previous expression cannot be computed by the estimator because the real timing vector  $\tau$  is not a priori known. Nevertheless, the gradient expected value close to the real timing vector does not depend on the absolute timing error  $\tau - \hat{\tau}$ . Hence, an accurate bias estimation can be obtained if the estimated timing vector is used to compute (26):

$$\widehat{\text{Bias}}_i = -2\text{Re} \{ \text{Tr} \{ \mathbf{\Gamma} \mathbf{A}_{\hat{\tau}}^H (\mathbf{I} - \mathbf{A}_{\hat{\tau}} \mathbf{C}_{\hat{\tau}})^H \mathbf{D}_{i_{\hat{\tau}}} \} \} \quad (27)$$

As a result, an unbiased estimation of  $\tau$  vector can be obtained according to a modified gradient where the bias is subtracted:

$$g_{mcv_i}^{\text{unbiased}} = -2\text{Re} \{ \mathbf{r}^H (\mathbf{I} - \mathbf{A}\mathbf{C})^H \mathbf{D}_i \mathbf{C} \mathbf{r} \} - \widehat{\text{Bias}}_i \quad (28)$$

## 5. PERFORMANCE ANALYSIS

This section derives the Gaussian Unconditional Cramér-Rao Bound (*UCRB*) to compare it with the proposed *CML* and *MCV* multi-user delay estimators analyzing its performance. As it is shown in [3] the *UCRB* is a valid lower bound for the variance of any consistent estimator based on the data sample covariance matrix.

As derived in [10] the  $ij$ th Fisher Information Matrix (*FIM*) element can be obtained as:

$$\{FIM_u\}_{ij} = \text{Tr} \{ \mathbf{R}^{-1} \mathbf{R}_i \mathbf{R}^{-1} \mathbf{R}_j \} \quad (29)$$

where:

$$\begin{aligned} \mathbf{R} &= E \{ \mathbf{r} \mathbf{r}^H \} = \mathbf{A} \mathbf{\Gamma} \mathbf{A}^H + \sigma_w^2 \mathbf{I} \\ \mathbf{R}_i &= \frac{\partial}{\partial \tau_i} \mathbf{R} \end{aligned} \quad (30)$$

Focusing on our estimation problem, assuming that the noise power is *a priori* known (which is considered in the *MCV* case), and modeling the transmitted symbols to be zero mean independent random variables (i.e.  $\mathbf{\Gamma}$  is a diagonal matrix)  $\mathbf{R}_i$  results:

$$\begin{aligned} \mathbf{R}_i &= \sigma_w^2 (\mathbf{D}_i \mathbf{A}^H + \mathbf{A} \mathbf{D}_i^H) \quad i = 1 \dots N_s \\ \mathbf{D}_i &= \frac{\partial}{\partial \tau_i} \mathbf{A} \end{aligned} \quad (31)$$

which can be substituted into (29) to obtain the *UCRB*.

## 6. SIMULATION RESULTS

To evaluate the *CML* (15) and *MCV* (28) estimators its performance was compared computing the Root-Mean Square Error (*RMSE*) in the timing delay estimation, and the Bit-error Rate (*BER*) in the symbol detection. Simulations were done considering 5 users, the spreading codes were Gold sequences with 7 chips per bit, the pulse shaping was a square-root raised cosine pulse with roll-off factor equal to 0.5 and the considered modulation was BPSK, and the oversampling factor was  $N_{sc} = 2$ . Denoting  $B_L$  as the equivalent noise loop bandwidth, this parameter is related with the number of transmitted symbols as [12]:

$$B_L T = \frac{1}{2N_s} \quad (32)$$

and the *UCRB* lower bound is usually written as a function this bandwidth factor.

Figure 1 compares the proposed *MCV* versus the classical *CML* algorithm, and compares the *RMSE* with the derived *UCRB* lower bound assuming that the noise power  $\sigma_w^2$  is *a priori* known (29) - (31). A low *SNR* scenario with *near-far*  $NF=0$  and only one path per user on *AWGN* (i.e. no channel assumption) was simulated. As it can be seen in figure 1, due to the high eigenvalue spread, at low *SNR* the *CML* is not an optimal solution and does not achieve the derived *UCRB*. Under those conditions, the proposed *MCV* outperforms the *CML* algorithm and attains the *UCRB*, becoming a quadratic optimal solution. Figure 1 also illustrates how at high *SNR* the *MCV* becomes the *CML* solution, and asymptotically both attain the Cramér-Rao Bound.

A second simulation shows the performance of both algorithms in symbol detection, and illustrates once again the importance of *MCV* in noisy environments. Figure 2 compares the *BER* according to the ML estimation of vector  $\mathbf{x}$  (13) considered in *CML* estimation, and the MMSE estimator (17) introduced in the *MCV* approach. In order to illustrate the eigenvalues spread importance, two simulations, using 7 chips per bit spreading codes (associated eigenvalue spreading  $\chi : 35$ ) and 15 chips per bit spreading codes (associated eigenvalue spreading  $\chi : 6.25$ ), were done. As it can be seen, the higher the eigenvalue spreading is, the worse the *CML* solution performs. When the system

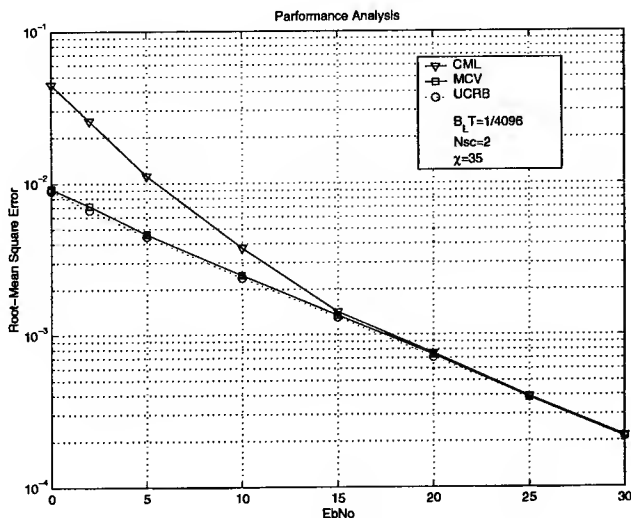


Figure 1: Timing Delay Estimation Error

is working at the limit of its capacity, (i.e. 5 users and spreading factor 7) the noise power is extremely increased by the decorrelating detector, and *CML* is not an acceptable solution, while the novel *MCV* always achieves a better performance.

## 7. CONCLUSIONS

In this paper the *MCV* algorithm has been introduced in the multiuser propagation delay estimation context. This novel method modifies the classical *CML* solution considering the impact of the noise in the Likelihood function compression. Hence, a more robust algorithm in noisy environments when the transference matrix eigenvalue dispersion is large, can be derived.

Simulations have shown *MCV* outperforms the classical deterministic algorithm in noisy conditions, and it corresponds asymptotically with the *CML* at high *SNR*'s. The mean squared timing error and the bit-error rate at the symbol detection have been used to evaluate this performance. Accordingly, the suggested quadratic estimation technique is shown to be optimal since it attains the *UCRB* lower bound in the whole *EbNo* range, becoming a great substitute not only to *CML*, but also to *UML* because it achieves similar features in a straightforward way.

## 8. REFERENCES

- [1] E.G. Ström, S. Parkvall, S.L. Miller, B.E. Ottersten, "Propagation delay estimation in asynchronous direct-sequence code-division multiple access systems", *IEEE Trans. on Commun.*, vol 44, Jan. 1996.
- [2] Z.Liu, J. Li, S.L. Miller, "An Efficient Code-Timing Estimator for Receiver Diversity DS-CDMA Systems", *IEEE Trans. on Communications*, vol 46, Jun. 1998.
- [3] P.Stoica, A.Nehorai. "Performance Study of Conditional and Unconditional Direction Of-Arrival Estima-

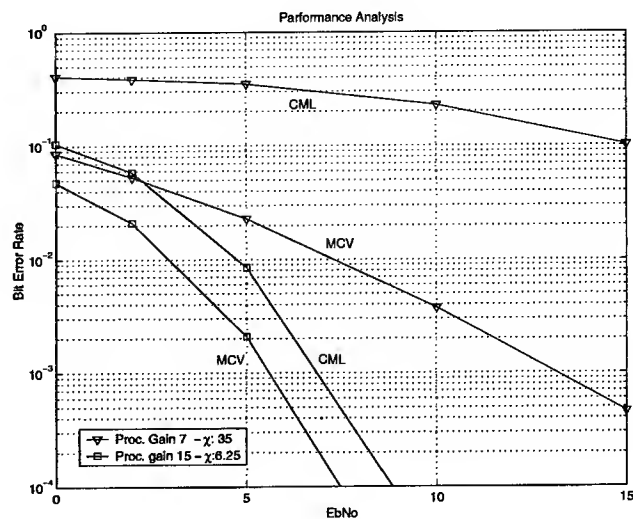


Figure 2: BER Symbol Detection

tion", *IEEE Trans. on Acoustics, Speech and Signal Processing*, vol.38, October 1990.

- [4] J.Riba, G. Vázquez, S. Calvo. "Conditional Maximum Likelihood Frequency Estimation for Staggered Modulations". *Proc. of ICASSP'98, Seattle (USA)*.
- [5] J.Riba, G.Vázquez. "CML Timing Recovery", *Proc. of ICASSP'99, Phoenix (USA)*.
- [6] J. Riba, J. Sala, G. Vázquez. "Conditional Maximum Likelihood Timing Recovery: Estimators and Bounds". Submitted to *IEEE. Trans on Communications* 1999.
- [7] G. Vázquez, J.Riba. *Non-Data-Aided Digital Synchronization*. In G. B. Giannakis, Y. Hua, P. Stoica, and L. Tong, editors. *Signal Processing Advances in Wireless Communications*, volume. II: Trends in Single and Multi-User Systems, chapter. 9. Prentice-Hall, 2000. To be published.
- [8] S.E. Bensley, B. Aazhang, "Subspace Based Channel Estimation for Code Division Multiple Access Communication Systems", *IEEE Trans. on Communications*, vol 44, August 1996.
- [9] M.Viberg, B. Ottersten, T.Kailath. "Detection and Estimation in Sensor Arrays Using Weighted Subspace Fitting". *IEEE Transactions on Signal Processing*, vol 39, November 1991.
- [10] B. Ottersten, M. Viberg, P. Stoica, *Radar Array Processing*. Springer-Verlag, 1993. Chapter: Exact and Large Sample ML Techniques for Parameter Estimation and Detection.
- [11] F. Rey, G. Vázquez, J. Riba. "Near-Far Resistant CML Propagation Delay Estimation and Multi-user Detection for Asynchronous DS-CDMA Systems". *Proc. of VTC-fall'99, Amsterdam (The Netherlands)*.
- [12] Umberto Mengali, Aldo N. D'Andrea, *Synchronization Techniques for Digital Receivers*, Plenum Press, 1997.

# NEW CRITERIA FOR BLIND EQUALIZATION OF $M$ -PSK SIGNALS

Zhengyuan Xu and Ping Liu

Dept. of Electrical Engineering  
University of California  
Riverside, CA 92521  
{dxu, pliu}@ee.ucr.edu

## ABSTRACT

Most digitally modulated signals can be depicted by signal space diagram. Such signal constellation often shows certain properties.  $M$ -PSK modulated signals uniformly lie on a circle. To equalize this class of signals, the constant modulus algorithm (CMA) is very efficient due to signals' constant modulus property. Its criterion penalizes deviations in the amplitude of equalized signals from a fixed value while ignoring the uniformly distributed phase. In this paper we explore both amplitude and phase properties in the equalization context. By combining dispersion of both the amplitude and phase value in one cost function, new criteria for blind equalization are obtained. Comparisons between the proposed methods and the CMA algorithms are made based on the level of inter-symbol interference and the probability of detection error.

## 1. INTRODUCTION

In digital communication, the user's information stream is usually modulated before transmission. Due to multipath propagation, inter-symbol interference (ISI) is introduced in the received signal. A equalizer is required to counter the effect of channel distortion. When source alphabets form a  $M$ -PSK constellation, the constant modulus algorithm (CMA) shows much efficiency in eliminating the interference due to other undesired symbols. An excellent review about this algorithm can be found in a recent paper [5]. After the algorithm was developed by [3] and [8], it has been extensively studied. The convergence property of CMA has been analyzed [2]. Connections between CMA and Wiener receivers are also built based on a novel geometrical concept [4]. It has been proved that the zero cost can be achieved by this criterion under some conditions [5].

As is well known, the constant modulus criterion employs the constant modulus (amplitude) property of modulated signals. In fact besides the property for the amplitude, most digitally modulated signals also

show other features such as uniform distribution on a plane in discrete intervals (quadrature amplitude modulation), on a unit circle (phase modulation) or on the real axis (amplitude modulation) in the signal space. These information will help equalize the channel, as employed in [1], [6] to match the signal constellation. Generally for a complex signal, it is described by its amplitude together with its phase. In this paper we focus on  $M$ -PSK modulated signals. Properties of other modulated signals can be similarly captured.

An  $M$ -PSK signal  $s$  can be represented by its constant amplitude  $r_0$  and uniformly distributed phase  $\Phi = \frac{2\pi m}{M}$  where  $m$  is a random number taking values  $0, 1, \dots, M-1$  with equal probability. It can be shown that the  $k$ -th order moment of this signal is zero for all  $k$  except that  $k$  is a multiple of  $M$  where it becomes a constant. Thus the moment instead of the absolute moment contains sufficient phase information. If we consider the  $M$ -th order moment of the equalized signal, then its norm square can be maximized to obtain the equalizer under certain constraint. This constrained maximization problem can also be converted into a constrained minimization problem based on our analysis.

In these approaches, the amplitude and phase of the equalized signal are jointly taken into account implicitly. However, explicit consideration is possible by integrating the phase property into the CMA cost function. It can be easily observed that the phase constraint  $\frac{M}{2}\Phi = m\pi$  is equivalent to  $\sin(\frac{M}{2}\Phi) = 0$ . Thus in order to achieve a better equalization performance, this phase deviation should also be minimized. Based on these observations, another criterion can then be developed by combining it with the CMA cost function. Similar to CMA algorithm, equalizers corresponding to different approaches can be recursively updated using the stochastic gradient ascent/descent method. Simulation results are presented to compare the proposed equalizer with CMA equalizer based on ISI and probability of detection error.



## 2. PROBLEM STATEMENT

Consider a widely adopted input/output model in wireless communications and blind equalization [4]

$$\mathbf{x}(n) = \mathbf{H}\mathbf{s}(n) + \mathbf{w}(n) \quad (1)$$

where  $\mathbf{s}(n) \in C^m$  is the complex source vector from  $M$ -PSK modulation constellation,  $\mathbf{H} \in C^{p \times m}$  is the channel matrix,  $\mathbf{w}(n) \in C^p$  represents additive white Gaussian noise (AWGN),  $\mathbf{x}(n) \in C^p$  is the received signal. The equalization is performed by designing an equalizer  $\mathbf{f} \in C^p$  whose output  $y_n$  is expected to be an accurate estimate of one of the elements in  $\mathbf{s}$

$$y_n = \mathbf{f}^H \mathbf{x}(n) = \mathbf{a}^T \mathbf{s}(n) + \mathbf{f}^H \mathbf{w}(n) \quad (2)$$

where  $(\cdot)^T$ ,  $(\cdot)^H$  stand for transpose and Hermitian,  $\mathbf{a}^T = \mathbf{f}^H \mathbf{H}$  is the combined response of the channel and the equalizer. Perfect equalization can be achieved in the absence of noise if  $\mathbf{a}$  has only one non-zero element [7]

$$\mathbf{a} = e^{j\theta} [0, \dots, 0, 1, 0, \dots, 0]^T \quad (3)$$

The position of the non-zero element in  $\mathbf{a}$  stands for the delay and  $\theta$  is the phase shift. Therefore the delay and phase ambiguity are inherent in blind equalization. Different criteria can be used to obtain the equalizer. The CMA criterion seeks to minimize the dispersion of the equalizer output about a constant  $r$

$$J_c(\mathbf{f}) \triangleq E\{(|y_n|^2 - r)^2\} \quad (4)$$

where "E" represents expectation. The constant can be chosen as  $r = \frac{E\{|s|^4\}}{E\{|s|^2\}}$  [3], [7]. For  $M$ -PSK signals,  $r = r_0^2$ . Due to high non-linearity of the cost function, the algorithm is usually implemented by stochastic gradient descent method

$$\mathbf{f}(k+1) = \mathbf{f}(k) - \mu(|y_k|^2 - r)y_k^* \mathbf{x}(k) \quad (5)$$

where  $*$  represents conjugate. It has been proved that zero cost can be achieved under some conditions on the source, channel and additive noise [5]. Since the phase characteristic of  $M$ -PSK signals is not captured in the CMA cost function, we will develop new criteria to jointly consider the properties of constant modulus and uniformly distributed phase values.

## 3. DEVELOPMENT OF THE CRITERIA

$M$ -PSK signals are uniformly distributed on a circle with radius  $r_0$ . Without loss of generality, we assume  $r_0 = 1$ . This constant modulus property together with

uniform phase values result in the following representation of the modulated signals

$$s = e^{j\Phi}, \quad \Phi = \frac{2\pi m}{M}, \quad m = 0, 1, \dots, M-1 \quad (6)$$

where  $m$  is a random number taking  $M$  possible values with equal probability  $\frac{1}{M}$ .  $M$  is usually chosen to be even  $M = 2L$ . By simple calculation, it can be verified that the  $k$ -th order moment of  $s$  satisfies

$$E\{s^k\} = \begin{cases} \frac{1 - e^{j2k\pi(M-1)}}{1 - e^{j\frac{2k\pi}{M}}} = 0 & k \neq lM \\ 1 & k = lM \end{cases} \quad (7)$$

Therefore based on (7) and the i.i.d. assumption of  $s_i$ , the  $M$ -th order moment of the equalizer output in the absence of noise is related to the combined impulse response by

$$\begin{aligned} E\{y_n^M\} &= E\left\{\left(\sum_l a_l s_{n-l}\right)^M\right\} \\ &= \sum_l a_l^M E\{s_{n-l}^M\} = \sum_l a_l^M \end{aligned} \quad (8)$$

with all cross terms zeroed out in the transition from the first line to the second line. Similarly we can obtain the output power [7]

$$\begin{aligned} E\{|y_n|^2\} &= E\left\{\left|\sum_l a_l s_{n-l}\right|^2\right\} \\ &= \sum_l |a_l|^2 E\{|s_n|^2\} = \sum_l |a_l|^2 \end{aligned} \quad (9)$$

Equations (8) and (9) form the basis to the following theorems.

**Theorem 1:** If  $E\{|y_n|^2\} = 1$ , then  $|E\{y_n^M\}|^2 \leq 1$ . The equality holds if and only if  $\mathbf{a}$  takes the form (3).

*Proof:* We apply the norm property first

$$|E\{y_n^M\}| = \left| \sum_l a_l^M \right| \leq \sum_l |a_l^M| = \sum_l (|a_l|^2)^L$$

For the above equality to hold,  $a_l^M$  should be non-negative numbers, or all  $a_l$  are zero except one is non-zero. Since for real numbers  $|a_l|^2$ , we have

$$\sum_l (|a_l|^2)^L \leq \left(\sum_l |a_l|^2\right)^L = (E\{|y_n|^2\})^L = 1$$

The equality holds if and only if  $|a_l| = 1$  for one  $l$  ( $l_0$ ) while  $|a_l| = 0$  for all the rest. Combining these conditions we have  $a_{l_0} = e^{j\theta}$ .  $\square$

**Theorem 2:** If  $|E\{y_n^M\}| = 1$ , then  $E\{|y_n|^2\} \geq 1$ . The equality holds if and only if  $\mathbf{a}$  takes the form (3).

*Proof:* The proof is similar to that in Theorem 1. The following can be easily verified

$$E\{|y_n|^2\} = \sum_l |a_l|^2 = \left[\left(\sum_l |a_l|^2\right)^L\right]^{\frac{1}{L}} \geq \left(\sum_l |a_l|^M\right)^{\frac{1}{L}}$$

$$\geq \left( \left| \sum_l a_l^M \right| \right)^{\frac{1}{L}} = \left( |E\{y_n^M\}| \right)^{\frac{1}{L}} = 1$$

The equality holds if and only if  $a_{l_0} = e^{j\theta}$  for a particular  $l_0$ .  $\square$

Theorem 1 and 2 suggest the following equalization criteria respectively.

### 3.1. The first criterion

According to Theorem 1, the equalizer  $\mathbf{f}$  can be obtained by

$$\max |E\{y_n^M\}|^2, \quad \text{subject to } E\{|y_n|^2\} = 1$$

After substituting  $y_n$  from (2) into the above, we can construct the Lagrange cost function for this constrained maximization problem

$$J_1(\mathbf{f}) = E\{(\mathbf{f}^H \mathbf{x})^M\} E\{(\mathbf{x}^H \mathbf{f})^M\} + \lambda_1 (\mathbf{f}^H \mathbf{R} \mathbf{f} - 1) \quad (10)$$

where  $\mathbf{R} = E\{\mathbf{x} \mathbf{x}^H\}$ ,  $\lambda_1$  is an unknown Lagrange multiplier. To seek a maximizer of  $J_1(\mathbf{f})$ , the gradient ascent method can be employed

$$\mathbf{f}(k+1) = \mathbf{f}(k) + \mu_1 \nabla J_1(\mathbf{f})|_{\mathbf{f}=\mathbf{f}(k)} \quad (11)$$

where  $\mu_1$  is the step size. From (10), the derivative with respect to  $\mathbf{f}^H$  can be shown to be

$$\nabla J_1(\mathbf{f}) = M E\{(\mathbf{f}^H \mathbf{x})^{M-1} \mathbf{x}\} E\{(\mathbf{x}^H \mathbf{f})^M\} + \lambda_1 \mathbf{R} \mathbf{f} \quad (12)$$

The optimal  $\lambda_1$  is the one which makes (12) zero under the constraint. By setting (12) equal to zero and applying our constraint, we can obtain  $\lambda_1$

$$\lambda_1 = -M E\{(\mathbf{f}^H \mathbf{x})^M\} E\{(\mathbf{x}^H \mathbf{f})^M\} \quad (13)$$

Therefore the derivative becomes

$$\nabla J_1(\mathbf{f}) = M b^* [E\{(\mathbf{f}^H \mathbf{x})^{M-1} \mathbf{x}\} - b \mathbf{R} \mathbf{f}] \quad (14)$$

where  $b = E\{(\mathbf{f}^H \mathbf{x})^M\}$ . Substituting (14) in (11) we obtain our recursion for the equalizer. To estimate expected values in (14) from the data, we save values for  $(\mathbf{f}^H \mathbf{x})^M$ ,  $(\mathbf{f}^H \mathbf{x})^{M-1} \mathbf{x}$  and  $\mathbf{x} \mathbf{x}^H$  at each iteration, and average them based on all of their values up to the current iteration. Simulation results show that this is a good approximation with less computations. As can be observed, complexity of this algorithm is about  $O((p+M)p)$ .

### 3.2. The second criterion

By examining Theorem 2, we can also obtain the equalizer  $\mathbf{f}$  based on

$$\min E\{|y_n|^2\}, \quad \text{subject to } |E\{y_n^M\}|^2 = 1$$

After considering (2), the Lagrange cost function can be built as

$$J_2(\mathbf{f}) = \mathbf{f}^H \mathbf{R} \mathbf{f} + \lambda_2 [E\{(\mathbf{f}^H \mathbf{x})^M\} E\{(\mathbf{x}^H \mathbf{f})^M\} - 1] \quad (15)$$

with a new multiplier  $\lambda_2$ . To minimize  $J_2(\mathbf{f})$ , we formulate the gradient descent recursion for the equalizer

$$\mathbf{f}(k+1) = \mathbf{f}(k) - \mu_2 \nabla J_2(\mathbf{f})|_{\mathbf{f}=\mathbf{f}(k)} \quad (16)$$

The derivative is easily computed from (15) as

$$\nabla J_2(\mathbf{f}) = \mathbf{R} \mathbf{f} + \lambda_2 M E\{(\mathbf{f}^H \mathbf{x})^{M-1} \mathbf{x}\} \quad (17)$$

Based on the constraint,  $\lambda_2$  can be similarly obtained

$$\lambda_2 = -\frac{\mathbf{f}^H \mathbf{R} \mathbf{f}}{M} \quad (18)$$

Therefore the derivative (17) becomes

$$\nabla J_2(\mathbf{f}) = \mathbf{R} \mathbf{f} - \mathbf{f}^H \mathbf{R} \mathbf{f} E\{(\mathbf{f}^H \mathbf{x})^{M-1} \mathbf{x}\} \quad (19)$$

Substituting (19) in (16) and using the technique as in the previous subsection to estimate expected values, we can finally obtain the recursion for the equalizer from data samples only.

The computational complexity of these two methods are very similar. It is significant if  $M$  is large. Next we will develop an alternative criterion, still based on the observed properties of  $M$ -PSK signals but with reduced complexity.

### 3.3. An alternative approach

Let us revisit the representation of  $s$  in (6). As far as the signal representation on the signal space is concerned, the phase property is equivalent to  $\sin(\frac{M}{2}\hat{\Phi}) = 0$ . Besides the constant modulus criterion, the deviation of  $\sin(\frac{M}{2}\hat{\Phi})$  from zero should thus be minimized as well, where  $\hat{\Phi}$  is the phase of the equalized signal  $y_n$ . Therefore we may construct the following cost function

$$J_3(\mathbf{f}) = E\{(|\mathbf{f}^H \mathbf{x}|^2 - 1)^2\} + \gamma E\{\sin^2(\frac{M}{2}\hat{\Phi})\} \quad (20)$$

where the first term is from  $J_c(\mathbf{f})$ ,  $\gamma$  is a weighting factor. By minimizing (20), the equalizer can be obtained. However, it is a highly non-linear function of  $\mathbf{f}$ . Similarly a gradient descent method has to be used

$$\mathbf{f}(k+1) = \mathbf{f}(k) - \mu_3 \nabla J_3(\mathbf{f})|_{\mathbf{f}=\mathbf{f}(k)} \quad (21)$$

The derivative of the first term on the RHS of (20) is

$$\mathbf{d}_1 = 2E\{(|y_k|^2 - 1)y_k^* \mathbf{x}\} \quad (22)$$

which is a function of  $\mathbf{f}$ . However, if we compute the derivative of the second term, we find that it requires the derivative of  $\hat{\Phi}$  (denoted by  $\hat{\Phi}_f$ )

$$\mathbf{d}_2 = \frac{M}{2} E\{\sin(M\hat{\Phi})\hat{\Phi}_f\}$$

To obtain its expression, we first define the real and imaginary parts of  $y_k$  by  $y_1$  and  $y_2$  respectively. Both of them can be expressed by  $\mathbf{f}$

$$y_k = y_1 + jy_2, \quad y_1 = \frac{\mathbf{f}^H \mathbf{x} + \mathbf{x}^H \mathbf{f}}{2}, \quad y_2 = \frac{\mathbf{f}^H \mathbf{x} - \mathbf{x}^H \mathbf{f}}{2j}$$

Then  $\hat{\Phi}$  is related to  $\mathbf{f}$  by

$$\hat{\Phi} = \arctan \frac{y_2}{y_1} = \arctan \frac{y_k - y_k^*}{j(y_k + y_k^*)} \quad (23)$$

Therefore  $\hat{\Phi}_f$  can be shown to be

$$\hat{\Phi}_f = \frac{\mathbf{x}^H \mathbf{f}}{2j|y_k|^2} \mathbf{x} = \frac{1}{2jy_k} \mathbf{x} \quad (24)$$

Hence  $\mathbf{d}_2$  becomes

$$\mathbf{d}_2 = \frac{M}{4j} E\left\{\frac{\sin(M\hat{\Phi})}{y_k} \mathbf{x}\right\} \quad (25)$$

Based on (22) and (25), finally we obtain  $\nabla J_3(\mathbf{f})$

$$\nabla J_3(\mathbf{f}) = E\{2(|y_k|^2 - 1)y_k^* \mathbf{x} + \frac{M \sin(M\hat{\Phi})}{4jy_k} \mathbf{x}\} \quad (26)$$

where  $\hat{\Phi}$  is given by (23) and  $y_k$  by (2). Substituting (26) in (21) and using instantaneous approximation for the expected values, we obtain the recursion for the equalizer

$$\mathbf{f}(k+1) = \mathbf{f}(k) - \mu_3 c \mathbf{x} \quad (27)$$

where

$$c = \frac{8j(|y_k|^4 - |y_k|^2) + M \sin(M\hat{\Phi})}{4jy_k}$$

(27) shows that at each iteration the equalizer is adjusted by the scaled data vector. This algorithm has complexity about  $O(p)$ .

#### 4. SIMULATIONS

We test the proposed methods and make comparisons with two typical CMA algorithms [3] [7] by computer simulations. Due to lack of space and the similarity between the first and the second proposed criteria, we only present the simulation results of the first criterion

and alternative approach for blind equalization of M-PSK signals. Two different measures will be adopted. First, inter-symbol interference (ISI) is used to demonstrate the convergence of the algorithm

$$ISI = \frac{\sum_i |\mathbf{a}_i|^2 - |\mathbf{a}|_{max}^2}{|\mathbf{a}|_{max}^2}$$

where  $\mathbf{a}^T = \mathbf{f}^H \mathbf{H}$ ,  $|\mathbf{a}|_{max}$  is the maximal absolute value of all elements in  $\mathbf{a}$ . Clearly, when  $\mathbf{a}$  has only one nonzero component as in (3),  $ISI = 0$  which is the ideal situation. Small ISI indicates the proximity to the desired response. Secondly, the probability of decoding error is especially meaningful in the communications context and also serves as an indicator of convergence. It is obtained from multiple independent realizations with random input signals and defined as the percentage of accumulated decoding errors among total number of transmitted symbols up to the current iteration.

In the experiments, we consider an unknown non-minimum phase channel used in [7] with unit sample response truncated at  $i = 3$  as: 0 when  $i < 0$ ,  $-0.4$  when  $i = 0$ , and  $0.84 \times 0.4^{i-1}$  when  $i > 0$ . Inputs are 4-PSK signal source with 4 equiprobable values: 1,  $-1$ ,  $+j$ ,  $-j$ . The step size  $\mu$  is set to 0.005. We use a 12-tap equalizer with the initial value as  $[0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0]^T$ . The iteration number is set to 5000. All the experiment results are obtained from 50 independent realizations.

In the first experiment, we compare the proposed first criterion with Shalvi's approach [7]. The first 400 iterations are based on [7] to obtain good initialization for both methods. The average ISI is plotted in Fig. 1 after 400 data points, where the solid line represents the proposed method while the dashed line for Shalvi's method. As expected, the proposed method converges to a much lower ISI level while maintaining almost the same fast convergence. Similar result can be observed from Fig. 2 for the error probability. The second experiment compares the proposed alternative approach with the CMA algorithm [3].  $\gamma$  is chosen to be 0.5. The average ISI and error probability are plotted in Fig. 3 and Fig. 4 respectively. Solid lines represent the proposed method while dashed lines for CMA. It can be observed that the proposed method converges faster than the standard CMA while achieving a lower ISI level after convergence. The error probability of the proposed method is also much lower than CMA.

#### 5. REFERENCES

- [1] S. Barbarossa and A. Scaglione, "Blind equalization using cost function matched to the signal constellation", *Proc. of 31st Asilomar Conference on*

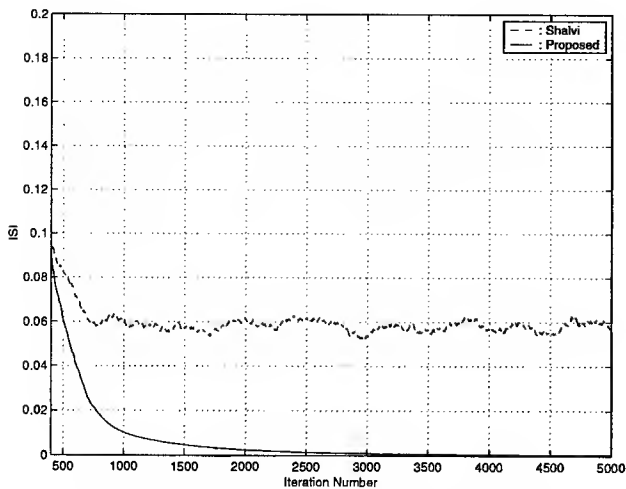


Figure 1: ISI of the first criterion and Shalvi's method.

*Signals, Systems and Computers*, vol.1, pp.550-4, Pacific Grove, CA, Nov. 1997.

- [2] Z. Ding, "On convergence analysis of fractionally spaced adaptive blind equalizers", *IEEE Trans. on Signal Processing*, vol. 45, pp. 650-657, Mar. 1997.
- [3] D.N. Godard, "Self-recovering equalization and carrier tracking in two dimensional data communication systems", *IEEE Trans. on Comm.*, vol. 28, no. 11, pp. 1167-75, November 1980.
- [4] M. Gu and L. Tong, "Geometrical Characterizations of Constant Modulus Receivers", *IEEE Trans. on Signal Processing*, vol. 47, no. 10, pp. 2745-2756, October 1999.
- [5] C.R. Johnson, *et.al*, "Blind Equalization Using Constant Modulus Criterion: A Review", *Proc. of the IEEE*, vol. 86, no. 10, pp. 1927-1950, October 1998.
- [6] Ta-Hsin Li and K. Mbarek, "A blind equalizer for nonstationary discrete-valued signals", *IEEE Transactions on Signal Processing*, vol.45, no.1, pp.247-54, Jan. 1997.
- [7] O. Shalvi and E. Weinstein, "New criteria for blind deconvolution of nonminimum phase systems (channels)", *IEEE Transactions on Information Theory*, vol.36, no.2, pp.312-21, March 1990.
- [8] J.R. Treichler and B.G. Agee, "A new approach to multipath correction of constant modulus signals", *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-31, pp. 459-472, April 1983.

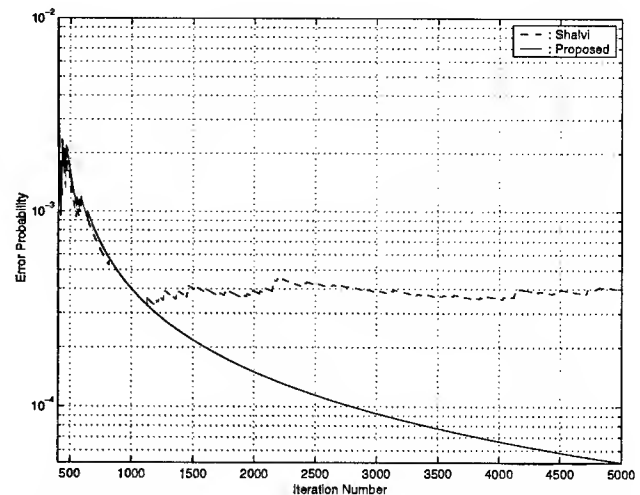


Figure 2: Error probability of the first criterion and Shalvi's method.

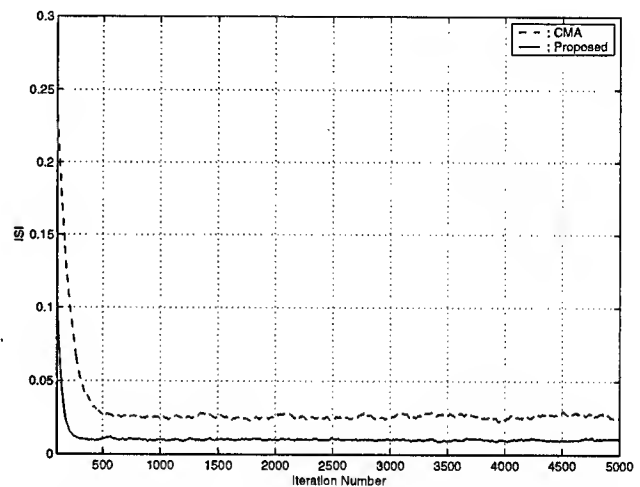


Figure 3: ISI of the alternative method and Godard's method.

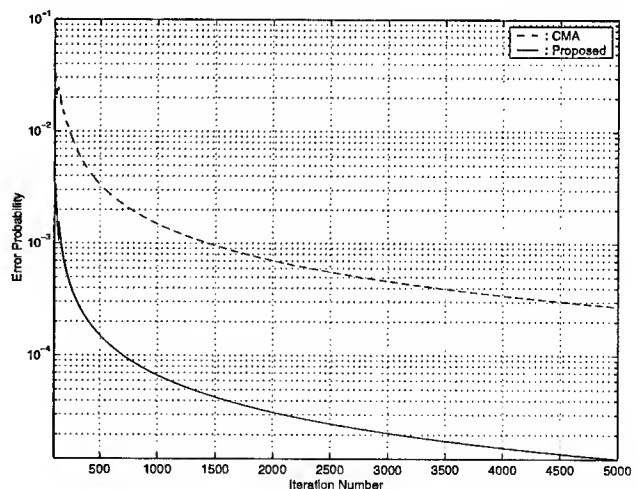


Figure 4: Error probability of the alternative method and Godard's method.

# THIRD-ORDER BLIND EQUALIZATION PROPERTIES OF HEXAGONAL CONSTELLATIONS

Charles D. Murphy

Signal Processing Lab  
Helsinki University of Technology  
P.O. Box 3000  
FIN-02015 HUT, FINLAND  
cmurphy@wooster.hut.fi

## ABSTRACT

Many digital communications systems use symbol constellations such as  $M$ -ary PAM, PSK, or QAM, which are simple to implement and symmetric. In a blind setting, equalization of inter-symbol interference (ISI) caused by a linear channel with unknown phase characteristics requires banded statistics of at least fourth order. Disadvantages of fourth-order statistical approaches are the relatively large sample sizes needed for good estimates of the statistics and the high computational complexity of the associated equalization algorithms. Here we examine several inherently asymmetric constellations with symbols placed on a hexagonal lattice rather than on the square lattice of QAM-type constellations. These constellations have a smaller average power under minimum symbol separation constraints than several widely-used constellations. We show how to modify the constellations to provide controllable third-order properties, and demonstrate blind equalization using a low-complexity third-order algorithm.

## 1. INTRODUCTION

Most constellations used in digital communications are symmetric. Well-known symmetric constellations include  $M$ -ary PAM, PSK, and square or cross-shaped QAM, along with a variety of specialized QAM-type constellations [1], [2].

A constellation  $\mathcal{S}$  from which a random symbol sequence  $\{x[n]\}$  is generated is *symmetric* when there is at least one rotation of the constellation by an angle  $\theta \in (0, 2\pi)$  that preserves all of the symbol values and their probabilities. An immediate difficulty when using a symmetric constellation is that it is not possible to derive a blind absolute reference phase at the receiver. This problem is ordinarily overcome via differential encoding of the data at the transmitter. A second problem is that the third-order statistics of linear combinations of the random symbols - the third-order statistics of linear channel outputs - do not contain information about the amplitude and phase properties of the channel. If the channel is known *a priori* to be a minimum-phase channel, or to be a maximum-phase channel, then second-order statistics suffice to equalize inter-symbol interference (ISI). If, however, such knowledge of the channel phase is absent, then statistics of at least fourth-order

are required to determine the channel phase. Fourth-order statistical algorithms are undesirable because they tend to have a high computational complexity and because accurate estimates of fourth-order statistics may require large sample sets.

In this paper, we investigate the statistical properties of "optimum" hexagonal constellations [3], [4], which are fairly old in concept but which were quickly dropped in practice in favor of the symmetric constellations already mentioned. It turns out that some of these hexagonal constellations are naturally asymmetric. This asymmetry allows extraction of channel phase (and amplitude) characteristics from third-order statistics. Moreover, the symbol separation of these hexagonal constellations is somewhat greater than that of their common symmetric counterparts given an average transmitted power constraint. In a sense, then, the asymmetry is "free", because it may be obtained without sacrificing immunity to additive noise.

On the design front, we modify the hexagonal constellations to increase the level of asymmetry while maintaining a minimum symbol separation and staying within an average transmitted power limit. The controlled third-order statistical characteristics of the constellations demonstrate that asymmetry can be a valuable tool in creating new types of symbol constellations that are resistant to noise, ISI, and other distortions. Finally, we demonstrate blind equalization of ISI-corrupted channels using the asymmetric hexagonal constellations and a third-order algorithm.

## 2. SYSTEM MODEL

The end-to-end communications system we consider is digital, but models both analog and digital effects. The transmitted symbols  $\{x[n]\}$  from the constellation  $\mathcal{S}$  form an i.i.d. random sequence. The receiver knows the values and probabilities of the symbols in  $\mathcal{S}$ , but never with certainty any  $x[n]$ .

$$y[n] = \sum_{k=k_1}^{k_2} h_k x[n-k] + w[n] \quad (1)$$

The distorted channel outputs are modeled by (1), with ISI represented by the finite set of time-invariant channel taps,

$h_{k_1}, \dots, h_{-1}, h_1, \dots, h_{k_2}$  and additive white Gaussian noise  $w[n]$ .

The goal of the receiver is to estimate the transmitted symbol  $\hat{x}[n]$  for each  $n$ , ideally equal to  $x[n]$  with high probability. If the channel taps are known, the estimate may be generated using maximum-likelihood sequence estimation (MLSE), a feed-forward linear filter, or a decision-feedback structure [1]. Equalization in this paper will take the form of tap identification.

$$y_{mix}[n] = -0.5e^{-j\frac{\pi}{4}}x[n+1] + (1 + 0.45e^{-j\frac{7\pi}{12}})x[n] - 0.9e^{-j\frac{\pi}{3}}x[n-1] + w[n] \quad (2)$$

It is well known that second-order baud-sampled statistics of channel outputs do not contain separable information about the phase properties of the channel. The channel in (2) is a mixed-phase channel with zeros at  $2e^{-j\frac{\pi}{4}}$  and  $0.9e^{-j\frac{\pi}{3}}$ . The frequency response appears in Fig. 1. The amplitude is identical to that of a minimum-phase channel with zeros at  $0.5e^{j\frac{\pi}{4}}$  and  $0.9e^{-j\frac{\pi}{3}}$ , but the phase response is much different. An MMSE inverse filter for one of the channels will restore approximately flat amplitude and linear phase to that channel, but will fail to equalize the phase distortion of the other. The challenge for blind algorithms is to equalize both the amplitude and phase.

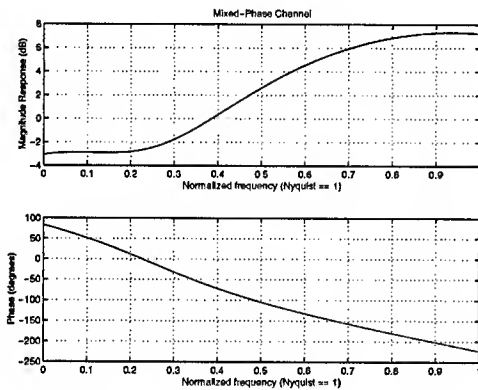


Figure 1: Frequency Response of (2)

Bussgang filters form a class of blind equalization algorithms that use feed-forward linear filters and low-complexity tap update equations. However, Bussgang filters are plagued by convergence problems of ability to find and speed of finding good filter tap values.

An approach with guaranteed convergence to globally-optimum parameter sets is to employ algorithms which use higher order statistics (HOS), such as the tricepstrum equalization algorithm (TEA) [5]. HOS algorithms typically exploit properties of fourth-order cumulants. They can be set up to produce inverse filter taps or estimates of the channel taps. However, obtaining reliable estimates of fourth-order cumulants may require a large set of channel outputs, and the computational complexity of HOS algorithms can be very high.

Third-order statistics are a viable alternative to fourth-order statistics, provided that they yield the same information about the channel phase. Advantages that third-order

algorithms may offer over fourth-order algorithms include reduced complexity and a shorter acquisition time for the statistical estimates. Third-order statistical phase information is not present in symmetric constellations, so we turn instead to asymmetric constellations.

### 3. HEXAGONAL CONSTELLATIONS

We are interested in  $M$ -ary constellations, where  $M = 2^N$  for some integer  $N > 0$ .  $N$  data bits are mapped to each successive symbol, a convenient feature from a design perspective. Classical examples of such constellations that we will consider are 8-PSK and 16-QAM. We will compare these to asymmetric "optimal" hexagonal constellations having 8 and 16 symbols [2]. For purposes of comparison, all of the constellation measurements will assume unit minimum symbol spacing unless otherwise specified.

The hexagonal constellations first appeared as a solution to the problem of maximizing the distance between symbols while minimizing the average transmitted power [3]. A few symmetric hexagonal constellations saw deployment in individual products, but the rest remained mostly a curiosity, unheralded and unwanted.

#### 3.1. 8-Point Constellations

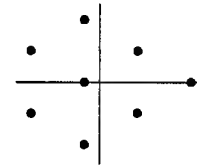


Figure 2: 8-HEX, An Asymmetric Hexagonal Constellation

The 2-point hexagonal constellation is identical to the usual binary constellation. The 4-point hexagonal constellation has a characteristic diamond shape. It has a lower average power and a higher peak power than the 4-QAM (4-PSK) constellation, but is not useful for our purposes because it is symmetric. The first suitable asymmetric hexagonal constellation has 8 symbol points as shown in Fig. 2. We refer to it hereafter as 8-HEX. Note that for any particular orientation of the constellation in the complex plane, there is a unique set of symbol values.

A key third-order statistic is  $\gamma_3$ , given in (3).

$$\gamma_3 = E\{|x[n]|^2 x[n]\} \quad (3)$$

The value of this moment is a measure of how much asymmetry a constellation exhibits. For the 8-HEX constellation with unit-spaced symbols,  $\gamma_3 = 0.1015$ . In contrast, for 8-PSK - a symmetric constellation -  $\gamma_3$  is identically zero.

The average power of the 8-HEX constellation,  $P_{average}$ , is 1.0781. The average power of 8-PSK with a minimum symbol spacing of 1 is 1.7071. The 8-HEX offers a 2 dB improvement over the average power of 8-PSK. With respect to blind equalization, the important improvement is the non-zero  $\gamma_3$ .

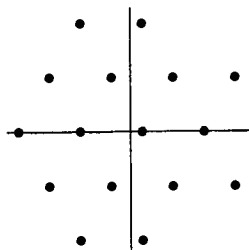
In some systems, particularly those in which amplifiers have a limited range of linearity, the peak power,  $P_{peak}$ ,

Constellation	$\gamma_3$	$P_{average}$	$P_{peak}$
8-PSK	0	1.7071	1.7071
8-HEX	0.1015	1.0781	2.2969
16-QAM	0	2.5	4.5
16-HEX	0.0938	2.1875	3.8125

Table 1: Constellation Comparisons Given Unit Minimum Symbol Separation

cannot be too large.  $P_{peak}$  of the 8-HEX constellation is 2.2969, as opposed to 1.7071 for 8-PSK. The increased  $P_{peak}$  of 8-HEX may or may not represent a desired characteristic.

### 3.2. 16-Point Constellations



The 16-point optimal hexagonal constellation (16-HEX) appears in Fig. 3.2. It is currently the best known 16-point configuration for maximizing the symbol spacing subject to an average transmitted power limit [2]. A commonly-used symmetric constellation having 16 points is 16-QAM, with the symbols occupying points on a four-by-four square lattice.

The value of  $\gamma_3$  for 16-HEX is 0.0938, less than the  $\gamma_3$  of 8-HEX, but greater than that of 16-QAM, which is 0. The average powers of the two 16-point constellations are 2.1875 for 16-HEX and 2.5 for 16-QAM. 16-HEX has a 0.6 dB average power advantage over 16-QAM, an approximate gain which is retained by higher-order  $2^N$ -ary hexagonal constellations over fully-filled square  $2^N$ -ary QAM [4].

The peak power of 16-QAM is 4.5, while the peak power of 16-HEX is only 3.8125. 16-HEX has lower  $P_{average}$  and  $P_{peak}$  than 16-QAM, and also a non-zero  $\gamma_3$ .

Table 1 summarizes the properties of the constellations we have discussed. As the numbers of points in the hexagonal constellations increase, the relative amount of asymmetry as measured by  $\gamma_3$  decreases, while  $P_{peak}$  and  $P_{average}$  are less than the corresponding quantities of the QAM constellations. These trends are a result of the sphere-packing nature of the hexagonal constellations: as more lattice points are available inside a complex-plane circle of a given diameter, it becomes easier to find a roughly-uniform, roughly-circular distribution.

## 4. MODIFIED HEXAGONAL CONSTELLATIONS

Hexagonal constellations we have not presented here that are of interest to digital communications systems designers include those with  $M = 32, 64, 128, 256, 512,$  and  $1024$

points, all of which are potential replacements for high-order  $M$ -ary QAM constellations in severely bandwidth-limited channels. A symmetric 64-point optimum hexagonal constellation was shown in [4].

For both the present optimum hexagonal constellations and the high-order varieties, there is room to increase the value of  $\gamma_3$  and thus the amount of asymmetry contained in the constellation. Having a larger  $\gamma_3$  is useful because the third-order statistics to be used for blind equalization will have a lower relative variance and the blind equalization algorithm can perform well more quickly than when  $\gamma_3$  is small.

A modified hexagonal constellation with the same minimum symbol spacing and the same  $P_{average}$  as the corresponding QAM constellation might replace the QAM constellation, provided that other specifications such as a maximum  $P_{peak}$  are not violated. We presently investigate changes to the optimum hexagonal constellations that involve moving one or two symbol points in the direction of the  $\gamma_3$  value, and the remaining symbol points in the opposite direction.

### 4.1. Modified 8-HEX

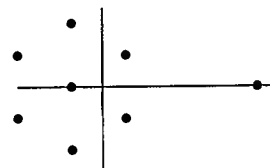


Figure 3: 8-HEX-MA, A Modified Asymmetric Hexagonal Constellation

If we add 1.0728 to the highest-power symbol in the 8-HEX constellation and restore the zero-DC condition by subtracting  $1.0728/7$  from each of the other symbols, we have the constellation 8-HEX-MA shown in Fig. 3. The average power of this constellation is 1.7071, identical to that of 8-PSK, while  $P_{peak} = 6.699$  and  $\gamma_3 = 1.5668$ .

The reason we choose this scheme of moving the highest-power symbol(s) in the direction of  $\gamma_3$  and the rest in the opposite direction is that  $\gamma_3$  is a power-weighted average. We try to greatly increase  $|x|^2 x$  ( $x \in S$ ) for one of the symbols while allowing  $|x|^2 x$  for the others to remain roughly the same.

### 4.2. Modified 16-HEX

In Fig. 4, we present a modified version of 16-HEX with increased  $\gamma_3$ . The two highest-power points have each been moved a distance of 0.5413 from their former locations, while the other 14 points were moved  $0.5413/7$  in the opposite direction. The 16-HEX-MA constellation has  $P_{average} = 2.5$ ,  $P_{peak} = 6$ , and  $\gamma_3 = 0.7365$ .

A second method one might use to increase the  $\gamma_3$  value of the 16-HEX constellation would be to pick one of the two highest-power symbols. Move it in the direction opposite that of a  $\gamma_3$  statistic excluding the selected point (i.e. using the other 15 symbols each with probability  $1/15$ ).

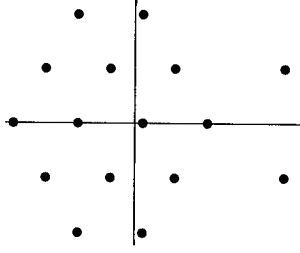


Figure 4: 16-HEX-MA, A Modified Asymmetric Hexagonal Constellation

Constellation	$\gamma_3$	$P_{average}$	$P_{peak}$
8-PSK	0	1.7071	1.7071
8-HEX-MA	1.0728	1.7071	6.6997
16-QAM	0	2.5	4.5
16-HEX-MA	0.5413	2.5	6

Table 2: Constellation Comparisons Given Unit Minimum Symbol Separation and Fixed  $P_{average}$ .

The statistical properties of the modified hexagonal constellations are shown in Table 2. Quite clearly, there is considerable leeway for constructing asymmetric constellations within the  $P_{average}$  and minimum symbol separation constraints imposed on typical systems where QAM constellations are in use. It is possible that this freedom may be parlayed into an "optimum" constellation having a minimum  $P_{average}$  subject to minimum symbol separation, minimum required  $\gamma_3$ , and maximum  $P_{peak}$  constraints.

## 5. EXAMPLES OF BLIND CHANNEL IDENTIFICATION

The natural asymmetry of some of the optimum hexagonal constellations and the strengthened asymmetry of the modified constellations are quite useful in blind equalization of linear channels modelled by (1). Consider the channel output statistic  $F_m = E\{|y[n]|^2 y[n+m]\}$ , which is related to  $\gamma_3$  of the constellation by (4).

$$E\{|y[n]|^2 y[n+m]\} = \gamma_3 \sum_{k=k_1}^{k_2} |h_k|^2 h_{k+m} \quad (4)$$

Since the channel length is finite, a time-domain method of estimating the channel taps can proceed in a simple, recursive fashion. Suppose that the channel length is  $L = k_2 - k_1 + 1$ .

$$F_{L-1} = \gamma_3 |h_{k_1}|^2 h_{k_2} \quad (5)$$

$$F_{-(L-1)} = \gamma_3 |h_{k_2}|^2 h_{k_1} \quad (6)$$

Equations (5) and (6) together permit estimation of the end-most channel taps  $h_{k_1}$  and  $h_{k_2}$ .

$$F_{L-2} = \gamma_3 |h_{k_1}|^2 h_{k_2-1} + \gamma_3 |h_{k_1+1}|^2 h_{k_2} \quad (7)$$

$$F_{-(L-2)} = \gamma_3 |h_{k_2-1}|^2 h_{k_1} + \gamma_3 |h_{k_2}|^2 h_{k_1+1} \quad (8)$$

Subsequent to obtaining  $h_{k_1}$  and  $h_{k_2}$ , solving (7) and (8) provides estimates of  $h_{k_1+1}$  and  $h_{k_2-1}$ . The remaining taps may all be computed in a similar fashion.

In reality, the receiver does not know the length of the channel, and must estimate  $\hat{L}$  based on statistics of the channel outputs. These estimated statistics - which may or may not be the  $\hat{F}_m$  values to be used in recovering the channel taps - and the  $\hat{F}_m$  estimates should converge with relative rapidity to the desired values. For  $\hat{F}_m$  estimates, the quality of the estimates can be improved by having a large  $\gamma_3$  rather than a small one, and by using larger channel output sets.

The channel for each of the simulations is (2), with AWGN having variance  $\sigma_w^2 = 0.1$ . For each of the asymmetric constellations, we generated 100 random sequences of 10000 symbols each, and estimated the channel taps using  $\hat{F}_m = \frac{1}{9990} \sum_{i=4}^{9993} |y[n]|^2 y[n+m]$  for  $m \in \{0, \pm 1\}$ . To clarify the performance and potential performance of the third-order approach to blind equalization for which we advocate asymmetric constellations, we plot the magnitude and phase of all the resulting channel estimates.

### 5.1. 8-HEX and 8-HEX-MA

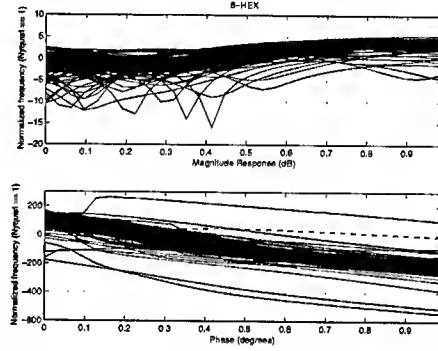


Figure 5: Channel Estimates Using 8-HEX

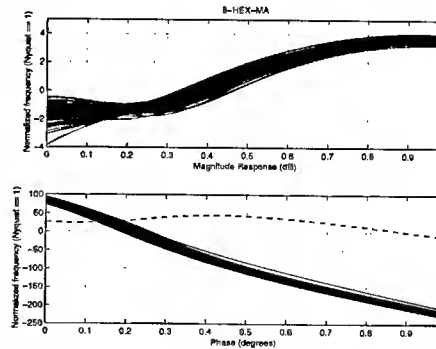


Figure 6: Channel Estimates Using 8-HEX-MA

In Figs. 5 and 6 we present the simulation results when the constellations are 8-HEX and 8-HEX-MA respectively. We can see in the first figure that several of the channel estimates based on the  $\gamma_3$ -weighted statistics of the 8-HEX constellation had fairly large amplitude and phase errors. On the other hand, the 8-HEX-MA, with its higher value of  $\gamma_3$ , provided very good estimates of both channel magnitude



and channel phase during each trial run. For reference, the phase response of the minimum-phase channel with the same amplitude response as (2) appears as a dashed line.

Because the channel estimates for the 10000-point data set using 8-HEX-MA were so good, we tried using smaller blocks. At 500 points, the algorithm still works well, while for blocks of 100 points the rate of poor channel estimates is appreciable - 25 to 30 out of 100 trials. However, the fact that it is possible to obtain reasonable channel estimates with fairly small data sets is a good sign and justifies our consideration of asymmetry in constellation design.

## 5.2. 16-HEX and 16-HEX-MA

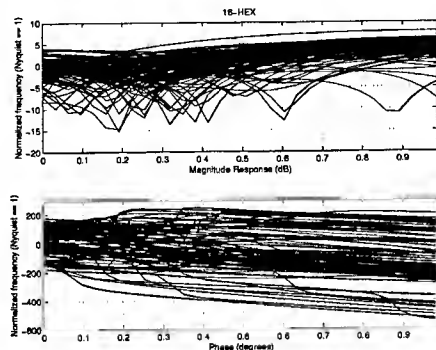


Figure 7: Channel Estimates Using 16-HEX

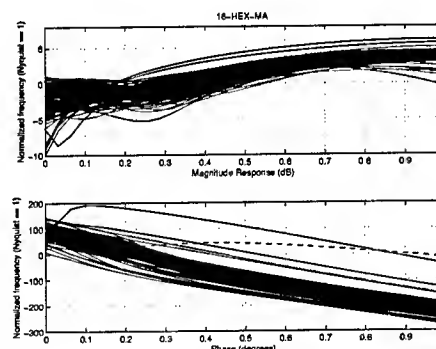


Figure 8: Channel Estimates Using 16-HEX-MA

The 100 channel estimates for the 16-HEX and 16-HEX-MA constellations appear in Figs. 7 and 8 respectively. Though there are many poor channel estimates for 16-HEX, when the added asymmetry is included, the 16-HEX-MA constellation allows the channel estimation algorithm to perform quite well.

## 5.3. Methods for Improving Performance

Ultimately, a blind equalizer does not operate as a standalone system. Rather, it is part of a larger whole which operates for a time using blind update and then switches to DD update. In this situation, even a relatively poor channel

estimate may provide the information needed to initialize (roughly) the equalizer parameter values.

Several avenues are available for improving the performance of the blind channel estimation. One we have discussed and shown is increasing  $\gamma_3$  for a given constellation. Another approach is to allow larger numbers of channel outputs to be used in the statistical estimates. If this is undesirable, then it might be possible to take advantage of other statistics than those we chose. For instance, second-order statistics contain information about tap amplitude. This information could be used to check the  $h_k$  values produced by our third-order algorithm to determine whether or not the estimated filter taps have appropriate amplitudes. Other third-order statistics contain amplitude and phase information which might be used in conjunction with that of the  $F_m$  statistics used in this paper.

## 6. CONCLUSION

We investigated the statistical properties of several venerable constellations that have not seen much use in actual communications systems. It turns out that some of these hexagonal constellations are asymmetric, so that information about the phase properties of the channel through which symbols from the constellation are transmitted is contained in third-order statistics of channel outputs. This enables use of third-order statistics to identify and equalize mixed-phase channels, in lieu of the fourth-order statistics which are required when the constellation is symmetric. This capability comes at low cost, as the hexagonal constellations have greater symbol separation than the popular symmetric constellations having the same average transmitted power. We showed a simple modification of the hexagonal constellations to increase their asymmetry while retaining minimum symbol separation and a limit on the average transmitted power, and showed channel equalization based on third-order statistics and the asymmetric constellations.

## REFERENCES

- [1] J.G. Proakis, *Digital Communications*, New York: McGraw-Hill 1995.
- [2] R.D. Gitlin, J.F. Hayes, and S.B. Weinstein, *Data Communications Principles*, New York: Plenum Press 1992.
- [3] G. Foschini, R.D. Gitlin, and S.B. Weinstein, "Optimization of Two-Dimensional Signal Constellations in the Presence of Gaussian Noise," *IEEE Trans. on Communications*, Vol. COM-21, No. 13, pp. 28-38, January 1974.
- [4] G.D. Forney Jr., R.G. Gallager, G.R. Lang, F.M. Longstaff, and S.U. Qureshi, "Efficient Modulation for Bandlimited Channels", *IEEE Journal on Selected Areas in Communications*, Vol. SAC-2, No. 5, pp. 632-647, September 1984.
- [5] D. Hatzinakos and C.L. Nikias, "Blind Equalization Based on Higher-Order Statistics (H.O.S.)", *Blind Deconvolution*, S. Haykin (ed.), PTR Prentice Hall, Englewood Cliffs, 1994.

# Comparison of the Cyclostationary and the Bilinear Approaches: Theoretical Aspects and Applications to Industrial Signals.

L. Bouillaut, M. Sidahmed

Heudiasyc - UMR CNRS 6599  
Université de Technologie de Compiègne  
BP 20529, 60205 Compiègne, FRANCE

## ABSTRACT

The aim of our study is to compare two a priori different approaches : Bilinearity and Cyclostationarity. Indeed, we underline that cyclostationary and bilinear tools make it possible to determine both non linear and non stationary links. These results are based on calculations, simulations for synthetic signals and finally applications to industrial signals. Then, we introduce different applications of these approaches to industrial vibrations. These methods enable us to obtain much more interesting results than classical methods such as Fourier, Spectrum, time-frequency studies... Finally we conclude on the interest and the reliability of such approaches and we introduce a method to determine if a link between two frequencies is rather cyclostationary than bilinear. This method is based on the use of higher-order cyclic statistics.

## 1. INTRODUCTION

The omnipresence of gears in most industrials sectors, as well as the need for an early diagnosis of faults and a quality control of noises, has made the study of vibrations a very interesting and exciting subject for the scientific world. Until now, vibration analysis was mainly based on stationary methods such as spectral analysis, Fourier analysis, cepstrum... [1], [2] and [3]. However, recent studies have proved that new methods based on non stationary and non linear properties of vibrating phenomena could bring results of the highest interest [4], [5] [6] [7].

Our study first deals with a comparison between cyclostationarity and bilinearity, and then presents different applications of these approaches.

In a first time, after a short introduction of main definitions of these approaches, we introduce a comparison of bispectrum and spectral correlation. Indeed, we underlined by calculation that cyclostationary and bilinear tools make it possible to determine both non linear and non stationary links [8]. Then, same results were presented by simulations on different synthetics signals [9] [10]. In this article, we present an application of this result to industrial signals, recorded on an helicopter of the NAVY (Westland Data).

Moreover, Cyclostationary and Bilinear approaches enabled us to obtain much more interesting results than classical methods such as Fourier, Spectrum analysis, time-frequency studies... Indeed, cyclic analysis allows a good early diagnosis [11]. We will present an application of the spectral correlation to

helicopter gearboxes vibrations. We will introduce the influence of torque on the quality of the diagnosis.

Finally, we introduce a first solution to determine if a link is more bilinear or more cyclostationary. This method is based on Higher-Order Cyclic Statistics, introduced by Garner [12] [13], and could allow to determine if a link between two frequencies is rather more cyclostationary or bilinear and, therefore, to determine if vibrations are more due to a surface fault or to a profile fault.

## 2. COMPARISON OF BILINEAR AND CYCLOSTATIONARY APPROACHES

### 2.1 Main definitions and properties

Contrary to a stationary signal, the cross correlation  $R_x(t, \tau)$  of which is only a function of  $\tau$ , a signal is said to be cyclostationary of second order when its cross correlation depends on the range variable  $\tau$  but is also periodically time dependent. The double Fourier transform of the cross correlation provides the spectral correlation  $S_x^\alpha(f)$  defined by :

$$S_x^\alpha(f) = FT_{t,\tau} [R_x(t, \tau)] = E \left[ X(f + \frac{\alpha}{2}) X^*(f - \frac{\alpha}{2}) \right] \quad (1)$$

The interpretation of results brought by the spectral correlation will be an interpretation in terms of statistically linked frequencies. Let's consider the random signal  $x(t) = a(t).e^{2j\pi.f_1 t} + b(t).e^{2j\pi.f_2 t}$  with  $a(t)$  and  $b(t)$  two random and stationary modulations. The expression of the spectral correlation of this signal, the calculations of which are presented in [7], led to the following conclusions:

If  $a(t)$  and  $b(t)$  are non correlated, the only non nil terms of  $R_x(t; \tau)$  are time independent. So, the signal is stationary.

On the other hand, if  $a(t)$  and  $b(t)$  are statistically linked, the cross correlation becomes time dependent. Moreover, for the spectral frequency  $(f_1 + f_2)/2$ , there are two cyclic frequencies,  $f_2 - f_1$  and  $f_1 - f_2$ , for which the spectral correlation is non zero.

To conclude, for the spectral correlation, the appearance of a peak for the couple of frequencies  $(f; \alpha)$  means that the two frequency components  $f + \alpha/2$  and  $f - \alpha/2$  are statistically linked [11].

As the  $n^{\text{th}}$  order stationarity study is linked with  $n^{\text{th}}$  order

moment, the study of linearity is linked with  $n^{\text{th}}$  order cumulants and their  $n$  Fourier transforms :  $n^{\text{th}}$  order polyspectra. We will limit our study to second order non linearity, described by their Bispectrum :

$$B(\lambda_1; \lambda_2) = \sum_{-\infty}^{+\infty} \sum_{-\infty}^{+\infty} C_3(\tau_1; \tau_2) e^{-2j\pi(\lambda_1\tau_1 + \lambda_2\tau_2)} = FT_{\tau_1, \tau_2} [C_3(\tau_1; \tau_2)] \quad (2)$$

The common properties of cumulants and bispectrum (linearity, symmetry, invariance by phase shifting) are precisely introduced in [14]. We will only underline the main property of bispectrum: Quadratic phase coupling (QPC) detection. If we consider two signals  $x_1$  and  $x_2$ , given by:

$$\begin{aligned} x_1(t) &= e^{2\pi f_1 t + \phi_1} + e^{2\pi f_2 t + \phi_2} + e^{2\pi f_3 t + \phi_3} \\ x_2(t) &= e^{2\pi f_1 t + \phi_1} + e^{2\pi f_2 t + \phi_2} + e^{2\pi f_3 t + (\phi_1 + \phi_2)} \end{aligned} \quad (3)$$

with  $f_3 = f_1 + f_2$ , and  $\phi_1, \phi_2, \phi_3$  random and independent variables. The study of the PSD will only present the three frequency components  $f_1, f_2$  and  $f_3$  without any information concerning phases. On the contrary, bispectra of  $x_1$  and  $x_2$  are different : The bispectrum of  $x_1$  will be identically nil whereas the bispectrum of  $x_2$  will present a peak for the couple of frequencies  $(f_1; f_2)$  and all its symmetries. To conclude, the appearance in a bispectrum of a peak for the couple of frequencies  $(f_i; f_j)$  will underline a bilinear coupling of the frequency components  $f_i$  and  $f_j$ .

Finally, the last property that we want to present in this section concerns the detection of bilinear links and cyclostationary links by both bispectrum and spectral correlation. Indeed, we proved by calculation [9] and by simulations with synthetic signals [8] that bilinear approach and cyclic analysis could detect both bilinear and cyclostationary links. This result is quite interesting. Indeed, the estimation of bispectra is very long and must be realized for all the frequency domain. On the contrary, the calculation of the spectral correlation can be done for a single cyclic frequency and is very fast computation ; even if it requires long data. To conclude, it could be more interesting to study certain bilinear phenomena using a cyclostationary approach rather than a bilinear one.

## 2.2 Detection of bilinearity and Cyclostationarity by Bispectrum and Spectral correlation: Application to industrial vibrations

In this section, we get interested in the extension of the previously enounced property to industrial signals, recorded on an helicopter (Westland Data). The system will be more precisely presented in the next section. As we can see on figure 1, with the fault, two non linear links appear in the bispectrum. The first one characterizes a modulation phenomena between two meshing frequencies  $f_{m1}$  and  $f_{m2}$ . Indeed, we underlined that when a fault appears in our system, the meshing frequency  $f_{m1}$  modulated  $f_{m2}$  [10]. This phenomena shows a bilinear link between considered frequencies. Moreover, spectral analysis underlined the appearance of a frequency  $f_3$ , situated exactly between meshing frequencies  $f_{m1}$  and  $f_{m2}$ , which was not characteristic of any component of our system [15]. We proved in [9] and [10] that this frequency was the consequence of the link between meshing phenomena. The appearance of this frequency gives rise to a bilinear link between  $f_3$  and  $f_{m2}$ .

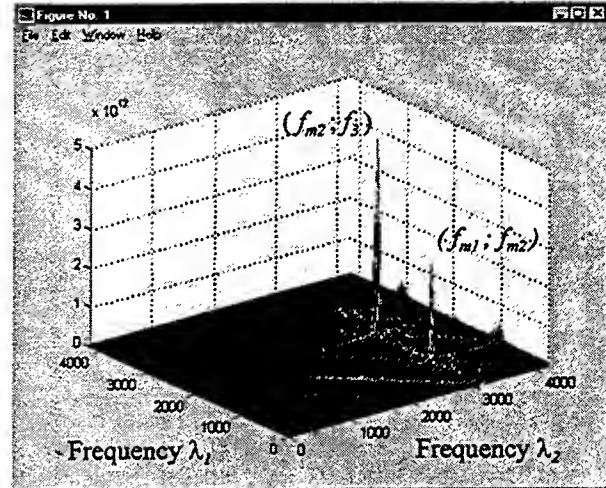


Figure 1. Bispectrum of Helicopter vibrations : Detection of bilinear links.

We wanted to know if we could encounter the result introduced in the section 2.1. result for industrial signals again. That is the reason why we estimated spectral correlations corresponding to eventual cyclostationary links between  $(f_{m1}; f_{m2})$  and  $(f_{m2}; f_3)$ . Figure 2 and Figure 3 present these estimations. As is shown in figure 2, during the research of eventual links between  $f_{m2}$  and  $f_3$ , the analysis of the spectral correlation underlines the appearance of two peaks, characterizing the existence of strong links between  $(f_{m1}; f_3)$  and  $(f_{m2}; f_3)$ .

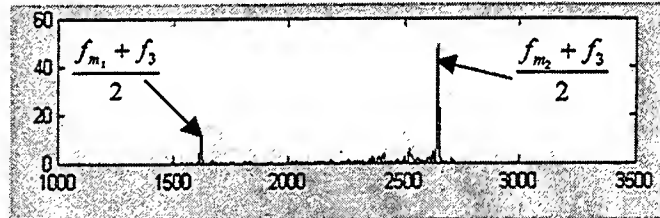


Figure 2. Spectral correlation for  $\alpha = f_{m2} - f_3 = f_3 - f_{m1}$ . Research of links between  $(f_{m1}; f_3)$  and  $(f_{m2}; f_3)$ .

These two peaks characterize strong links between  $f_3$  and meshing frequencies. The cyclic analysis therefore allows to detect the first peak of the bispectrum presented in figure 1. The second calculation concerns the research of eventual link between meshing frequencies. We therefore estimated the spectral correlation for  $\alpha = f_{m2} - f_{m1}$ .

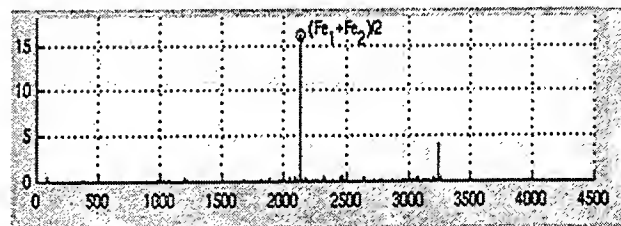


Figure 3. Spectral correlation for  $\alpha = f_{m2} - f_{m1}$ . Research of links between  $(f_{m1}; f_{m2})$ .

We can note that, for this cyclic frequency, a very significant peak appears for the spectral frequency  $f = (f_{m2} - f_{m1}) / 2$ . Thus, according to the property of interpretation in terms of statistically linked frequencies enounced previously, meshing frequencies are strongly correlated.

To conclude, this study confirm theoretical results obtain in [8] and [9] ; i.e. bispectrum and spectral correlation make it possible to detect and to determine precisely both non linear and non stationary links. This result raises a new problem. Indeed, the detection of a link using a bilinear approach or a cyclic analysis don't allow us to come to a conclusion about the nature of this link. However, such an information could be very interesting in terms of analysis of damage phenomena. Indeed, we know that according to the fault we are facing with (surface fault or profile fault), vibrations created by this fault will be rather more cyclostationary or bilinear. Thus, if detecting a link between frequencies (characterizing the apparition of a fault) we could conclude to the nature of this link, we could, in the mean time, determine if we are facing rather a spalling, a crack or a pitting... This wish motivated the study presented in section 4.

### 3. INFLUENCE OF TORQUE ON THE QUALITY OF THE DIAGNOSIS

#### 3.1 Introduction of the system

In this section, we are interested in the study of helicopter gearbox vibrations. These signals come from a measurement campaign realized by Westland Helicopter LTD on a NAVY equipment. Figure 4 presents a general view of our system. The principal component of our study is the CH46 gear box. Vibrations were recorded for eight different faults and eight different torque. For this paper, we will only study one spalling. Other faults have already been presented in [8] and [9].

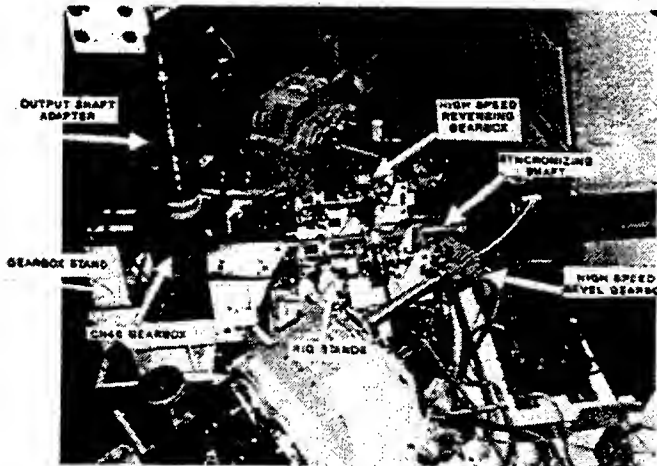


Figure 4. Photo of the helicopter engine.

The application of spectral correlation we want to present in this paper concerns the diagnosis of faults for helicopter vibrations. We know that the appearance of a fault for rotating machines is characterized by modulation phenomena [11]. Indeed, with the fault, rotating frequencies  $f_r$  will modulate meshing frequencies  $f_m$ . So a spectral analysis of these signals will underline the

existence of lateral bands, width  $f_r$ , around each harmonic of the meshing [15]. As described in [11], the appearance in the spectral correlation of a strong link between the meshing frequency and its lateral bands will underline the existence of a fault on the component characterized by these frequencies. The next study will be based on this property. The Figure 5 presents a simplified idea of our system.

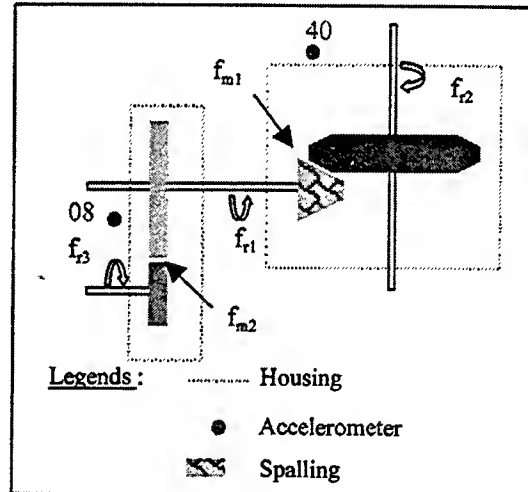


Figure 5. Simplified schema of our system.

We presented in [8], [9] and [10] that cyclic analysis offered very good result for early diagnosis of industrial systems (for simple systems as well as for complex systems). In the next section, we will present the influence of torque on the quality of this diagnosis.

#### 3.2 Influence of torque for diagnosis

The aim of the measurement campaign is to define different diagnostic processes, based on usual signal processing methods. The knowledge of the influence of the torque to diagnose faults could, therefore, allow us to contemplate an analysis, realized just before the helicopter takes off. So, inclining correctly paddles, it could be possible to realize a good quality diagnosis just before the use of the aircraft.

We, therefore, research links between meshing frequency  $f_m$  and its lateral bands to  $f_r$ . The figure 6 introduces the influence of the torque on the quality of the diagnosis. It represents the evolution of the characteristic peak corresponding to the link between the meshing frequency  $f_m$  and its lateral bands  $f_m + f_r$  for the different levels of torque.

First, it can be interesting to note that, for small faults, we will obtain a better diagnosis for the highest torque. This can be explained by the fact that the spalling of a tooth is the consequence of a weakness of the latter. So, the higher the torque will be, the stronger surface pressure will be. Thereby, the least change of surface will 'resound' stronger if the torque is high [16].

On the contrary, for an established fault, the influence of torque is completely inverted. The diagnosis will, therefore, be easier for a low level torque. Indeed, the previous small fault is now a

significant spalling. The fault becomes deeper and the meshing phenomenon appears as a shock. So, the higher the torque will be, the more the meshing teeth will be inclined to fit exactly the shape of the spalling ; decreasing, therefore, its repercussions in the vibrating signal [16].

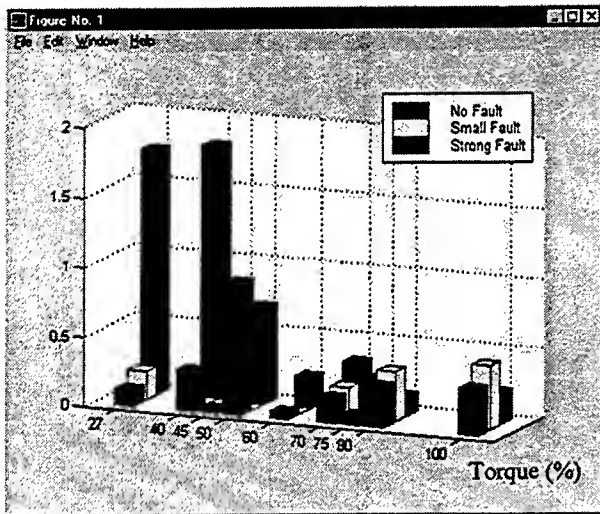


Figure 6. Spectral correlation : Evolution, with the torque, of the link between  $f_{m1}$  and  $f_{r1}$ .

Thus, it seems to be preferable to use low levels of torque to diagnose a relatively important fault. All these results were confirmed by a similar study on a crack propagation [8].

#### 4. CLASSIFICATION METHOD BASED ON HIGHER ORDER CYCLIC STATISTICS

As we note in section 2.2, bispectrum and spectral correlation detect both bilinear and cyclostationary links and thus, do not allow to determine the nature of such links. We, therefore, introduce Higher Order Cyclic Statistics as a possible solution of our problem [15] [16]. We will limit our study to second and third order cyclic statistics. Indeed, we proved that the use of cyclic bispectrum makes it possible to determine if a link between two frequency components is cyclostationary or bilinear. We will not present calculations of cyclic bispectrum in this paper. Indeed they are very long and will be completely introduced in a future paper in *Mechanical Systems and Signal Processing*, [18] and in the final report of my PhD.

The idea was to calculate cyclic bispectra of cyclostationary and bilinear synthetic signal  $x(t)$  and  $x_f(t)$ , described in section 2.1. Then we researched criteria allowing to distinguish bilinear link from cyclostationary correlations.

Figure 7 presents the complete cyclic bispectrum of the cyclostationary signal  $x(t)$ . We can note that, if amplitude modulations  $a(t)$  and  $b(t)$  are independent, the stationary version of  $x(t)$  is characterized by two peaks situated for the triplets of frequencies  $(\alpha; \lambda_1; \lambda_2) = (f_1; f_1; f_1)$  and  $(f_2; f_2; f_2)$ . On the contrary, if  $a(t)$  and  $b(t)$  are statistically linked, we can observe

that, for cyclic frequencies  $\alpha=f_1$  et  $\alpha=f_2$ , additional peaks appears for  $(f_1; f_2)$  et  $(f_2; f_1)$ . Moreover, two other peaks also characterized this cyclostationary link :  $(2f_2 - f_1; f_2; f_2)$  and  $(2f_1 - f_2; f_1; f_1)$ .

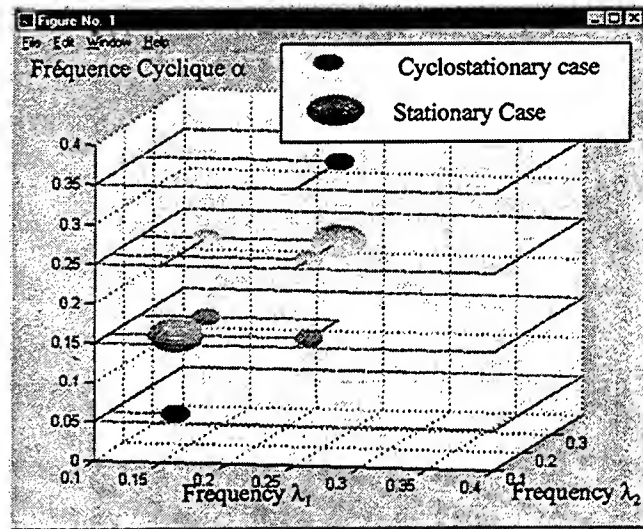


Figure 7. Cyclic Bispectrum of the more or less cyclostationary signal  $x(t)$ .

In a second time, the calculation of the theoretical cyclic bispectrum of the bilinear signal  $y(t)$  raises following cyclic frequencies [17] :

- For  $\alpha = f_1$ : Peaks appear for  $(\lambda_1; \lambda_2) = (f_1; f_1)$ ,  $(f_2; f_1)$  and  $(f_1; f_2)$  but also for  $(f_3; f_1)$  and  $(f_1; f_3)$ .
- For  $\alpha = f_2$ : We encounter peaks for  $(f_2; f_2)$ ,  $(f_2; f_1)$  and  $(f_1; f_2)$  again ; but also for  $(f_3; f_2)$  and  $(f_2; f_3)$ .
- For  $\alpha = f_3$ : Peaks appear for  $(f_3; f_3)$ ,  $(f_3; f_1)$  and  $(f_1; f_3)$  as also for  $(f_3; f_2)$  and  $(f_2; f_3)$ .
- For  $\alpha = 2f_1$ : Two peaks appear for the couple of frequencies  $(f_3; f_1)$  and  $(f_1; f_3)$ .
- For  $\alpha = 2f_2$ : Characteristic peaks of this frequency are situated in  $(f_2; f_3)$  and  $(f_3; f_2)$ .
- For  $\alpha = 2f_2 - f_1$ : A peak appears for  $(f_2; f_2)$ .
- For  $\alpha = 2f_1 - f_2$ : A peak appears for  $(f_1; f_1)$ .
- For  $\alpha = 2f_3 - f_1$ : Now, a peak appears for  $(f_3; f_3)$ .
- For  $\alpha = 2f_3 - f_2$ : The same peak appears for  $(f_3; f_3)$ .
- For  $\alpha = f_2 - f_1$ : We encounter a peak for  $(f_2; f_2)$  again.
- For  $\alpha = f_1 - f_2$ : The cyclic bispectrum is characterized by a peak for frequencies  $(f_1; f_1)$ .

These results allow us to set up a classification process to determine if a link between two frequencies  $f_1$  and  $f_2$  is rather more cyclostationary or bilinear. The following figure introduces the diagram of this method.

Nevertheless, the estimation of HOCS [13] still raises problems if we want to deal with industrial long data. This estimation can be based on a generalization of the estimation of Higher Order Statistics to non stationary processes. These methods commonly use periodograms. We are still working on this problem and we will present an application of the cyclic bispectrum to previously used helicopter signals in a future paper.



## 6. REFERENCES

- [1] Capdessus C. et Sidahmed M., *Analyse des vibrations d'un engrenage : Cepstre, Corrélation, Spectre, Traitement du signal*, Vol. 8, No. 5, pp. 365-372.
- [2] Sidahmed M. et Dus G., *Détection de défauts dans les engrenages par analyse vibratoire*, Congrès 'Engrenages et Transmissions Mécaniques', Paris, Février 1992.
- [3] Sidahmed M. et Garnier C., *Détection de défauts dans les engrenages*, CETIM Informations, Octobre 1991, pp.71-74.
- [4] Gardner W.A., *The spectral correlation theory of cyclostationary time-series*, 1986, No. 11, pp.13-36.
- [5] Gardner W.A., *Measurement of Spectral Correlation*, Trans. IEEE on Acoustic, speech, and Signal Processing, 1986, Vol. 34, No. 5, pp. 1111-1123.
- [6] Rubini R. and Sidahmed M., *Diagnostic of Gears Systems using the Spectral Correlation Density of Vibration Signal*, Congress SAFE Process '97, Hull, August 1997.
- [7] Bouillaut L. et Sidahmed M., *Approche Cyclostationnaire et Bilinéaire des signaux vibratoires d'engrenages*, 3<sup>rd</sup> International Conference Acoustical and Diagnostic Surveillance, Senlis, France, October 1998, pp. 323-332.
- [8] Bouillaut L. and Sidahmed M., *Cyclostationary an Bilinear Approaches for Gear Vibration Signals*, 9<sup>th</sup> IMEKO TC-10 International Conference on Technical Diagnostics, Wroclaw, Poland, 1999, pp.86-102.
- [9] Bouillaut L. and Sidahmed M., *Cyclostationary and Bilinear Approaches for gears vibrating signals*, Damage Assessment of Structures, Dublin, Ireland, 1999, pp.354-362.
- [10] Bouillaut L. and Sidahmed M., *Cyclostationary Approach for early diagnosis and physical analysis of industrial signals*, ISMA 25 International Conference on Noise and Vibration Engineering, Leuven, Belgium, August 2000.
- [11] Capdessus C., *Aide au diagnostic des machines tournantes par traitement du signal*, Thèse INPG, 1992.
- [12] Gardner W.A., *The Cumulant Theory of Cyclostationary Time-Series, Part I: Foundation*, IEEE Trans. On signal processing, Vol. 42, N°12, pp. 3387-3408, December 1994.
- [13] Gardner W.A., *The Cumulant Theory of Cyclostationary Time-Series, Part II: Development and Applications*, IEEE Trans. On signal processing, Vol. 42, N°12, pp. 3409-3429, December 1994.
- [14] Hinich M.J., *Higher Order cumulant and cumulant Spectra*, Circuit System and Signal Processing, Vol. 13, N°4, pp. 391, 1994.
- [15] Braun S., Sidahmed M., Feldman M. and Zacksenhouse M., *Vibration based gear diagnostic with application to Westland helicopter data*, 15<sup>th</sup> International Modal Analysis Conference, Orlando, February 8<sup>th</sup>-11<sup>th</sup> 1999.
- [16] El Badaoui M., *Contribution au diagnostic des réducteurs complexes à engrenages par l'analyse Cepstrale*, Thèse de l'université de Saint-Etienne-Jean-Monnet, 1999.
- [17] Bouillaut L. and Sidahmed M., *Comparaison des approches cyclostationnaire et bilinéaire : Aspect théorique et applications à des signaux réels*, Traitement du signal, publication in progress.

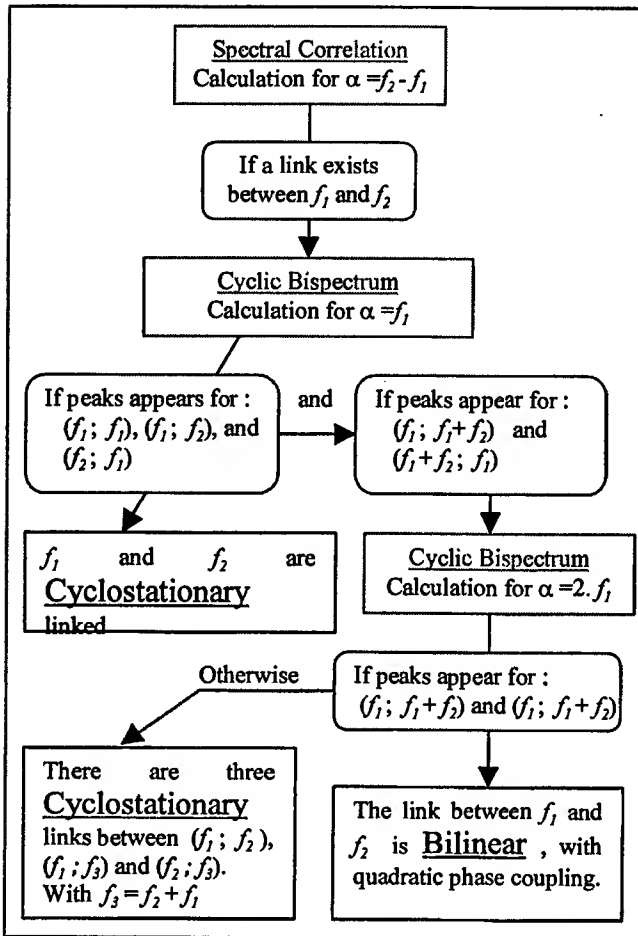


Figure 8. Cyclostationary or Bilinear classification method, based on the use of HOCS.

## 5. CONCLUSION

In this paper, we presented a comparison of bilinear approach and cyclostationary analysis. It appears that these approaches are closely linked. Indeed, bispectrum and spectral correlation make it possible to detect and to determine precisely both bilinear and cyclostationary links.

Thus, a cyclic analysis seems to be more appropriated to diagnose a fault on a system. Indeed, the estimation of the spectral correlation is a lot faster than bispectrum computation which requires an estimation over the complete frequency domain. Moreover, the cyclic approach makes it possible to perform a very good early diagnosis. Finally, we presented the influence of the torque on the quality of the diagnosis.

In the second part of this paper, we introduced the need of another approach, allowing to distinguish these two classes of signal. One solution was offered using higher order cyclic statistics.

However, for the moment, the problem of the estimation of HOCS for long industrial data is not completely solved and will be dealt with in a future study.

# ARRAY PROCESSING OF UNDERWATER ACOUSTIC SENSORS USING WEIGHTED FOURIER INTEGRAL METHOD

*I. S. D. Solomon and A. J. Knight*

Defence Science and Technology Organisation  
P.O.Box 1500, Salisbury 5108, Australia  
E-mail : daniel.solomon@dsto.defence.gov.au

## ABSTRACT

The spatial processing of platform-mounted acoustic sensors is complicated by platform generated noise. The Weighted Fourier Integral Method (WFIM) beamformer has been shown to perform well in such cases, by reducing this coloured noise which is received by the sensors. In this paper WFIM is modified by using maximum likelihood estimates of the spatial correlation lags. The proposed technique exploits the work of Burg et. al. and estimates these lags for a sparse redundant linear array of hydrophones. The results obtained illustrate the significant performance improvement obtainable over that of the least-squares lag estimation procedure utilised in WFIM. The proposed approach "better" estimates the contributions of missing sensors, in the sparse array, and performance approaching the full array, with extra sensors, is attained. A beamformer which adaptively weights the covariance lags is also proposed and preliminary results presented.

## 1. INTRODUCTION

The authors have been studying sonar beamforming of platform-mounted acoustic sensors; a problem that is complicated by platform generated noise which adversely effects the detection of sources (or "contacts") in a background of ambient noise. The authors have used a sparse linear array and have both analysed real data and conducted real-time at sea testing, to evaluate the performance of beamformers [1, 2]. Due to continuing advances, multi-processor digital systems are now capable of performing sophisticated signal processing in real-time [2]; hence more advanced techniques may now be considered for real-time applications.

The Fourier Integral Method (FIM) beamformer, developed by Wilson and Nuttall [3, 4] after a number of experiments at sea, was shown to outperform the conventional beamformer and was somewhat better than the Minimum Variance Distortionless Response (MVDR) beamformer. The authors have studied the

performance of FIM and a variant of it, called Weighted FIM (WFIM), which were suitably modified for sparse arrays. The results obtained with a platform-mounted array (see [1, 2]) showed that WFIM outperformed both FIM and MVDR. WFIM was able to significantly reduce the platform generated noise, which was unrejected by MVDR and the conventional beamformer (FIM did somewhat reduce this noise).

The Bartlett (or conventional) beamformer and the MVDR beamformer apply weights directly to the sensors; weighting is thus performed in the sensor domain. The WFIM beamformer applies weights in the spatial correlation lag domain. In the general case lag weighting can achieve everything that may be achieved in the sensor domain, and in addition can achieve more. Thus WFIM is capable of providing improvements over beamformers operating in the sensor domain.

The performance gains obtained by WFIM are due to two factors. Firstly, the lag weighting performed by WFIM reduces the contributions of small lags; this is where a significant amount of the noise (platform noise and also ambient noise) has been found to be present. By reducing the contributions of these lags, to the beamformer output, the adverse effects of noise are reduced and hence good performance is achieved by WFIM. The weighting used by WFIM also results in a narrow main beam (narrower than MVDR in practice) and reduced sidelobe levels.

The second reason for the improvements obtained by WFIM is the technique used for estimating the correlation lags. Using a least-squares approach, WFIM essentially compensates for the missing sensors in the sparse array. Hence performance similar to an array with extra sensors is achievable. In this paper, further improvements are considered which result in further reduction of energy leakage in sidelobe directions.

It should be noted that the power output of WFIM can go negative. As a result strong contacts can potentially cause the masking of weak contacts. In practice however the occurrence of this is not common, and this

drawback is out weighted by the improvements provided by WFIM for sonar beamforming applications. This issue is discussed later, in this paper, when adaptive weighting in the lag domain is considered.

## 2. ESTIMATION OF CORRELATION LAGS

Array signal processing of acoustic sources usually involves the computation of second order statistics of band-limited hydrophone data (see [1]). The second order statistics, for a particular frequency, are represented by the spatial covariance matrix

$$\mathbf{S} = \begin{bmatrix} s_{0,0} & s_{0,1} & \cdots & s_{0,M-1} \\ s_{1,0} & s_{1,1} & \cdots & s_{1,M-1} \\ \vdots & \vdots & \ddots & \vdots \\ s_{M-1,0} & s_{M-1,1} & \cdots & s_{M-1,M-1} \end{bmatrix} \quad (1)$$

where spatial covariance lag  $s_{m_1, m_2} = s_{m_2, m_1}^*$  represents the cross-correlation between sensors  $m_1$  and  $m_2$ .

The power output of the WFIM beamformer, as a function of bearing  $\theta$  and frequency  $f$ , is as follows [1]

$$p(\theta, f) = \frac{1}{2K-1} \sum_{k=-(K-1)}^{K-1} w_k \rho_k(f) v_k(\theta, f) \quad (2)$$

where  $K$  is the number of lags present,  $w_k$  are the lag weights,  $\rho_k(f)$  are the spatial correlation lags (at frequency  $f$ ) and  $v_k(\theta, f)$  is related to the array steering vector. The approach used in WFIM to estimate  $\rho_k$  (at a particular frequency) from  $\mathbf{S}$ , for a sparse linear array, is to employ a least-squares technique [1]. The array used by the authors was a redundant array, so the co-array is full and estimates of all correlations lags up to lag  $(K-1)$  are available.

The least-squares formulation used is to estimate a Hermitian covariance matrix which contains the necessary correlations lags [5, 6]. Note for a wide-sense spatially stationary process the covariance matrix is Toeplitz if the array is equispaced and the signals are uncorrelated; this has been exploited in [5, 6, 7] to obtain improved spatial processing. The least-squares technique provides an analytic solution which is close to the true covariance matrix in a minimum norm sense, but its optimality in practice for beamforming applications is questionable. In this paper an alternative approach is considered, which is to estimate the correlations such that the likelihood is maximised. The approach exploits the work by Burg et. al. in [8].

Burg et. al. considered the problem of spectral estimation given a uniformly sampled real time series. They considered the estimation of a covariance matrix of specified structure (usually Toeplitz), and did this

by maximising the likelihood function. Here the array processing problem is considered and, in particular, for sparse linear arrays with complex (Hermitian) covariance matrices. The approach by Burg et. al. can be modified for this problem and is now detailed.

The likelihood function can be expressed as

$$l(\rho) = -\ln(\det(\mathbf{R})) - \text{tr}\{\mathbf{R}^{-1}\mathbf{S}\} \quad (3)$$

where  $\mathbf{R}$  is a structured covariance matrix which is related to  $\rho = [\rho_0, \rho_1, \dots, \rho_{K-1}]^T$  which are to be estimated (note  $\rho_{-k} = \rho_k^*$  and  $\rho_k$ 's are for frequency  $f$ ).

To find the maximum of (3), one needs to differentiate  $l(\rho)$  with respect to the correlations  $\rho$  and set the resulting expression to zero. It can be shown [8] that

$$\frac{\partial l(\rho)}{\partial \rho} = \text{tr} \left\{ (\mathbf{R}^{-1}\mathbf{S}\mathbf{R}^{-1} - \mathbf{R}^{-1}) \frac{\partial \mathbf{R}}{\partial \rho} \right\} \quad (4)$$

which is set to equal zero for obtaining the maximum.

Burg et. al. realised that not only does  $\partial \mathbf{R} / \partial \rho$  satisfy equation (4) but also matrix  $\mathbf{Y}$  which is in the set of covariance matrices to be estimated, and so given a solution to the above equation  $\mathbf{R}$  (say an initial estimate) one needs to solve the following equation for  $\mathbf{R}'$  which is the new solution :

$$\text{tr}\{(\mathbf{R}^{-1}\mathbf{S}\mathbf{R}^{-1} - \mathbf{R}^{-1}\mathbf{R}'\mathbf{R}^{-1})\mathbf{Y}\} = 0 \quad (5)$$

or alternatively as

$$\text{tr}\{\mathbf{R}^{-1}\mathbf{S}\mathbf{R}^{-1}\mathbf{Y}\} = \text{tr}\{\mathbf{R}^{-1}\mathbf{R}'\mathbf{R}^{-1}\mathbf{Y}\} \quad (6)$$

Now one must write the matrix  $\mathbf{R}'$  in terms of the unknown correlations. The following expansion is employed for the problem considered here,

$$\mathbf{R}' = \rho_o \mathbf{I} + \sum_{k=1}^{K-1} (\rho_k^r \mathbf{X}_k^1 + \rho_k^i \mathbf{X}_k^2) \quad (7)$$

where  $\rho_k^r = \Re\{\rho_k\}$  and  $\rho_k^i = \Im\{\rho_k\}$  ( $\rho_o$  is real), the matrices  $\mathbf{X}_k^1$  and  $\mathbf{X}_k^2$  are sparse matrices which contain either  $\{1, +j, -j, 0\}$  in locations corresponding to  $\rho_k$  ( $j$  is the complex operator). With this expansion, equation (6) becomes

$$\begin{aligned} \text{tr}\{\mathbf{R}^{-1}\mathbf{S}\mathbf{R}^{-1}\mathbf{Y}\} &= \rho_o \text{tr}\{\mathbf{R}^{-1}\mathbf{R}^{-1}\mathbf{Y}\} + \\ &\sum_{k=1}^{K-1} \rho_k^r \text{tr}\{\mathbf{R}^{-1}\mathbf{X}_k^1\mathbf{R}^{-1}\mathbf{Y}\} + \\ &\sum_{k=1}^{K-1} \rho_k^i \text{tr}\{\mathbf{R}^{-1}\mathbf{X}_k^2\mathbf{R}^{-1}\mathbf{Y}\} \end{aligned} \quad (8)$$

which one can write in a similar form to [8], by noting  $\mathbf{Y} \in \{\mathbf{I}, \mathbf{X}_1^1, \dots, \mathbf{X}_{K-1}^1, \mathbf{X}_1^2, \dots, \mathbf{X}_{K-1}^2\}$  and defining

$$c_k = \text{tr}\{\mathbf{R}^{-1}\mathbf{S}\mathbf{R}^{-1}\mathbf{Y}\} \quad (9)$$



$$\mathbf{a}_k = [\text{tr}\{\mathbf{R}^{-1}\mathbf{R}^{-1}\mathbf{Y}\}, \text{tr}\{\mathbf{R}^{-1}\mathbf{X}_1^1\mathbf{R}^{-1}\mathbf{Y}\}, \dots, \text{tr}\{\mathbf{R}^{-1}\mathbf{X}_{K-1}^1\mathbf{R}^{-1}\mathbf{Y}\}, \text{tr}\{\mathbf{R}^{-1}\mathbf{X}_1^2\mathbf{R}^{-1}\mathbf{Y}\}, \dots, \text{tr}\{\mathbf{R}^{-1}\mathbf{X}_{K-1}^2\mathbf{R}^{-1}\mathbf{Y}\}]^T \quad (10)$$

and thus obtain a system of equations  $\mathbf{A}\mathbf{x} = \mathbf{c}$  as in [8], where the  $(2K-1) \times (2K-1)$  symmetric real matrix  $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_{2K-1}]$ , the  $(2K-1)$  element real vector  $\mathbf{c} = [c_k]$ , and the  $(2K-1)$  element real vector  $\mathbf{x} = [\rho_0, \rho_1^r, \dots, \rho_{K-1}^r, \rho_1^i, \dots, \rho_{K-1}^i]^T$ . Since some of the matrices above are sparse, direct application of the above equations make the algorithm computationally inefficient. By using the following, in equation (9) and (10), computation time is significantly reduced

$$\mathbf{X}_k^1 = \sum_{(l,n) \in \Omega_k} (\mathbf{E}_{l,n} + \mathbf{E}_{n,l}) \quad (11)$$

$$\mathbf{X}_k^2 = j \sum_{(l,n) \in \Omega_k} (\mathbf{E}_{l,n} - \mathbf{E}_{n,l}) \quad (12)$$

where  $\Omega_k$  is the set of matrix indices corresponding to lag  $k$  ( $\geq 0$ ),  $\mathbf{E}_{l,n} = \mathbf{e}_l \mathbf{e}_n^T$  ( $\mathbf{e}_l$  is the unit vector with all but the  $l$ th element zero). After some manipulation, one obtains  $\mathbf{c} = [c_1^0, c_1^1, \dots, c_{K-1}^1, c_1^2, \dots, c_{K-1}^2]^T$  and

$$\mathbf{A} = \begin{bmatrix} a_{0,0} & a_{0,1} & \dots & a_{0,K-1} \\ a_{1,0} & a_{1,1} & \dots & a_{1,K-1} \\ \vdots & \vdots & \vdots & \vdots \\ a_{K-1,0} & a_{K-1,1} & \dots & a_{K-1,K-1} \\ a_{K,0} & a_{1,1}^{2,1} & \dots & a_{1,K-1}^{2,1} \\ \vdots & \vdots & \vdots & \vdots \\ a_{2K-2,0} & a_{K-1,1}^{2,1} & \dots & a_{K-1,K-1}^{2,1} \\ a_{0,K} & \dots & a_{0,2K-2} \\ a_{1,1}^{1,2} & \dots & a_{1,K-1}^{1,2} \\ \vdots & \vdots & \vdots \\ a_{K-1,1}^{1,2} & \dots & a_{K-1,K-1}^{1,2} \\ a_{1,1}^{2,2} & \dots & a_{1,K-1}^{2,2} \\ \vdots & \vdots & \vdots \\ a_{K-1,1}^{2,2} & \dots & a_{K-1,K-1}^{2,2} \end{bmatrix} \quad (13)$$

where  $a_{l,m}$  are elements of  $\mathbf{A}$  as given by (10), and

$$c_1^0 = \text{tr}\{\mathbf{R}^{-1}\mathbf{S}\mathbf{R}^{-1}\} \quad (14)$$

$$c_k^1 = 2 \sum_{(l,n) \in \Omega_k} \Re\{(\mathbf{R}^{-1}\mathbf{S}\mathbf{R}^{-1})_{n,l}\} \quad (15)$$

$$c_k^2 = 2 \sum_{(l,n) \in \Omega_k} \Im\{(\mathbf{R}^{-1}\mathbf{S}\mathbf{R}^{-1})_{l,n}\} \quad (16)$$

$$a_{k,k'}^{1,1} = 2 \sum_{(l,n) \in \Omega_k} \sum_{(l',n') \in \Omega_{k'}} \Re\{(\mathbf{R}^{-1})_{n,l}(\mathbf{R}^{-1})_{n',l'} + (\mathbf{R}^{-1})_{n,n'}(\mathbf{R}^{-1})_{l',l}\} \quad (17)$$

$$a_{k',k}^{2,1} = 2 \sum_{(l,n) \in \Omega_k} \sum_{(l',n') \in \Omega_{k'}} \Im\{(\mathbf{R}^{-1})_{n',l}(\mathbf{R}^{-1})_{n,l'} - (\mathbf{R}^{-1})_{n',n}(\mathbf{R}^{-1})_{l,l'}\} \quad (18)$$

$$a_{k',k}^{2,1} = 2 \sum_{(l,n) \in \Omega_k} \sum_{(l',n') \in \Omega_{k'}} \Im\{(\mathbf{R}^{-1})_{n,l'}(\mathbf{R}^{-1})_{n',l} - (\mathbf{R}^{-1})_{n,n'}(\mathbf{R}^{-1})_{l',l}\} \quad (19)$$

$$a_{k,k'}^{2,2} = -2 \sum_{(l,n) \in \Omega_k} \sum_{(l',n') \in \Omega_{k'}} \Re\{(\mathbf{R}^{-1})_{n,l'}(\mathbf{R}^{-1})_{n',l} - (\mathbf{R}^{-1})_{n,n'}(\mathbf{R}^{-1})_{l',l}\} \quad (20)$$

Note  $a_{k,k'}^{1,2} = a_{k',k}^{2,1}$ , equations (18) is for  $k' \geq k$  and equation (19) is for  $k' \leq k$ .

The iterative procedure starts by using  $\mathbf{R} = \mathbf{I}$ , which is a positive definite matrix. Matrix  $\mathbf{A}$  and vector  $\mathbf{c}$  are then calculated using equations (14)-(20), and then vector  $\mathbf{x}$  is estimated. From the correlations lags, obtained from  $\mathbf{x}$ , the covariance matrix  $\mathbf{R}'$  is constructed using equation (7) and then one must check that the covariance matrix is both positive definite (i.e. all the eigenvalues are positive) and that it increases the likelihood (equation (3)); if not, try  $q\mathbf{R}' + (1-q)\mathbf{R}$  where  $q = 0.5$ . Continue halving  $q$  till the above conditions are met. Then set the new solution to  $\mathbf{R}$  and continue iterating until the likelihood does not increase much. Note the inverse of  $\mathbf{R}$  may be determined by exploiting the inherent structure present.

### 3. PERFORMANCE COMPARISON

The performance of the WFIM beamformer, using the correlation lags estimated above, is now considered by using them in equation (2). Figure 1 shows the power output of the Bartlett beamformer, the WFIM beamformer and the proposed WFIM-ML beamformer; the sparse array has half the sensors missing and the data had a strong contact present at most frequencies. The gray scales, which have the same dB range in each subfigure, have sufficiently large range to show the full extent of the improvements obtained using WFIM-ML. As can be seen WFIM reduces the platform generated noise, which is present in the conventional beamformer at low frequencies, and thus enhances the detection of contacts from the background. WFIM-ML is seen to similarly reduce the noise, but in addition is able to further reduce the leakage of the contact's energy in the sidelobe directions and was found to provide up to 10 dB improvement over WFIM.

The performance of WFIM-ML over WFIM are due to the improved estimation of the correlation lags. It was found that performance approaching that of a uniform linear array, with extra sensors, was achieved i.e. the results attained are similar to that theoretically

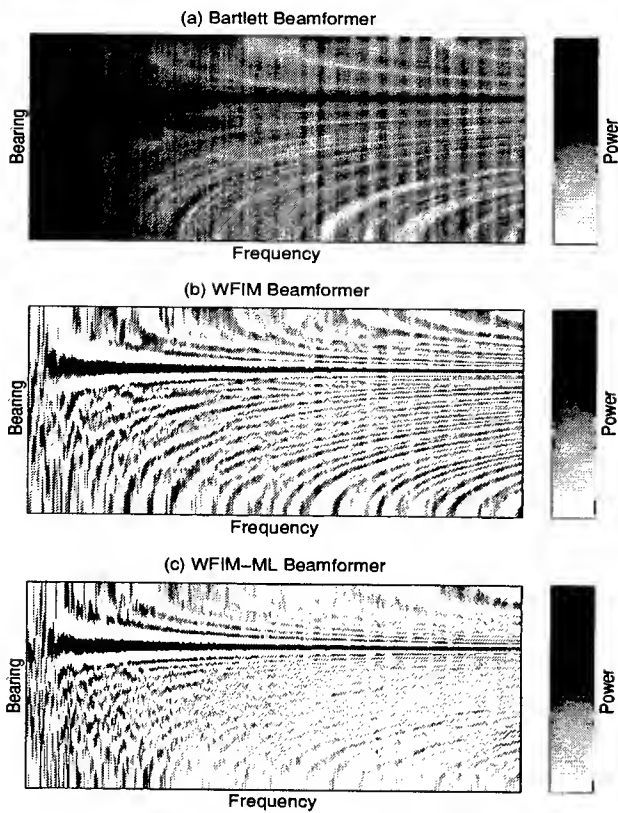


Figure 1: Beamformer Comparison.

achievable using a full array. Using the sparse array the spatial stationarity of the process has been exploited; reduced spatial sampling has been replaced by equivalent temporal sampling. It should be noted that in low-noise simulations the least-squares approach was found to give similar results to the new approach. The maximum likelihood correlation lag estimates may also be used in an augmented covariance matrix [5] to enhance the Bartlett beamformer; this is however not possible using a FFT based beamformer.

Figure 2 illustrates the performance of WFIM-ML when even more sensors are excluded, while still maintaining a full co-array. Figure 2(a) and 2(b) show the Bartlett beamformer and the WFIM-ML beamformer respectively, for the original sparse array; a strong contact is seen in both sub-figures, while WFIM-ML shows two additional weaker contacts. Figure 2(c), which was obtained when several more sensors were excluded, shows that the performance of WFIM-ML is comparable here to the original sparse array with some differences however observed. Figure 2(d) shows the same case as in 2(c) but here the number of snapshots, used to estimate the covariance matrix, has been doubled.

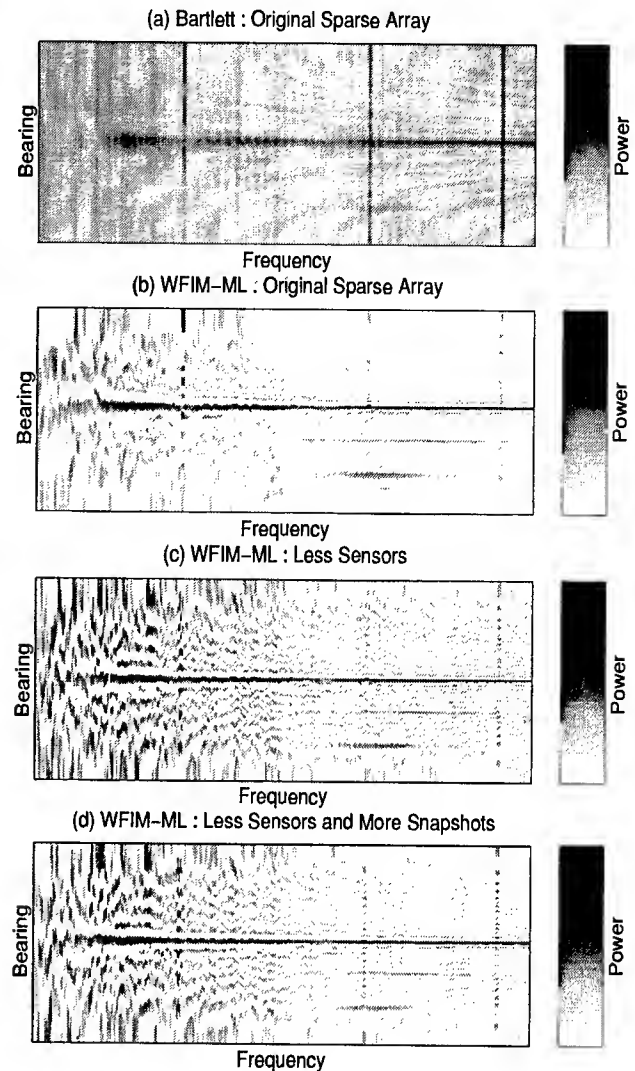


Figure 2: Performance with even less sensors.

The differences are seen to be reduced and indicates that the more snapshots obtained, from a stationary process, results in better performance. The algorithm was found to always converge and each iteration did provide observable improvements.

#### 4. ADAPTIVE FIM

The WFIM-ML results show the improvements that may be obtained by "better" combining the covariance lag contributions. Now a procedure developed for adaptively weighting the covariance lags is discussed. The algorithm minimises the power output of the beamformer while ensuring (a) the power output in the direction of interest is unity, and (b) the power output is

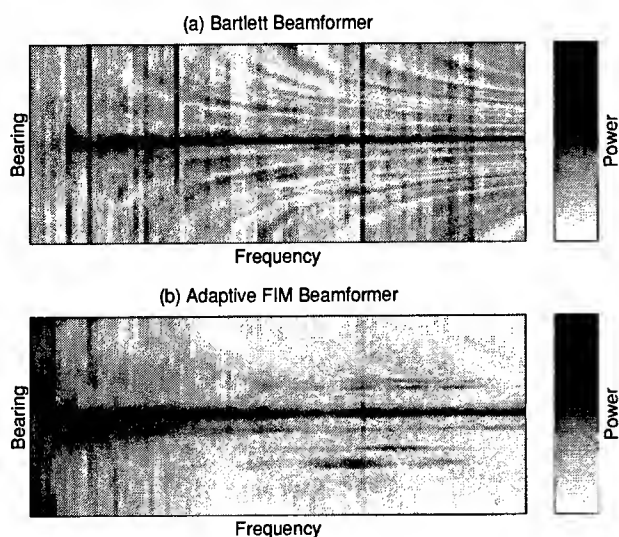


Figure 3: Adaptive lag-domain Beamformer.

always positive. When the covariance lags are weighted the power output is real provided Hermitian weights are used; the power output is however not guaranteed to be positive. Positive power output, or more generally beam pattern (filter response) positivity, is very important for adaptive covariance lag weighting, since otherwise the adaptivity can result in the nulling of contacts. Thus the adaptive algorithm is constrained such that the power output is always positive, by formulating the problem in a semidefinite optimisation framework [9]. The results obtained, for one such adaptive algorithm, is shown in figure 3. This beamformer is seen to extract several weak contacts which are not apparent in the Bartlett beamformer. Modifications, such as the controlled partial relaxation of constraint (b), are currently being investigated.

## 5. CONCLUSION

The WFIM beamformer, which has been previously shown to improve spatial processing of acoustic sources, has been further improved. A maximum likelihood procedure was used to estimate the correlation lags, which was shown to provide superior performance to the lag estimation procedure used in WFIM. The new WFIM-ML beamformer, like WFIM, performs well in the presence of platform generated noise, but WFIM-ML is seen to better detect contacts in the presence of a strong contact. The WFIM-ML beamformer's performance, with the sparse array considered, was seen to approach that attainable with a full array with ex-

tra sensors. An adaptive lag beamformer was also discussed, and it was shown to provide very promising early results. Further research is currently being conducted in this area and extensive testing of this adaptive beamformer is being performed.

## REFERENCES

- [1] I. S. D. Solomon, A. J. Knight, M. V. Greening, "Sonar array signal processing for sparse linear arrays", Fifth International Symposium on Signal Processing and its Applications (ISSPA 99), August 1999, Brisbane, Australia, Vol. 2. pp. 527-530.
- [2] I. S. D. Solomon, A. J. Knight, D. F. Liebing, S. B. Faulkner, "Sonar beamforming using Fourier Integral Method", Underwater Defence Technology (UDT) Conference, February 2000, Sydney, Australia, pp. 182-186.
- [3] J. H. Wilson, "Applications of inverse beamforming theory", Journal of the Acoustical Society of America, Vol. 98, No. 6, December 1995, pp. 3250-3261.
- [4] A. H. Nuttall, J. H. Wilson, "Estimation of the acoustic field directionality by use of planar and volumetric arrays via the Fourier Series Method and the Fourier Integral Method", Journal of the Acoustical Society of America, Vol. 90, No. 4, Part 1, October 1991, pp. 2004-2019.
- [5] S. U. Pillai, Y. Bar-ness, F. Haber, "A new approach to array geometry for improved spatial spectrum estimation", Proceedings of the IEEE, Vol. 73, No. 10, October 1985, pp. 1522-1524.
- [6] D. A. Gray, B. D. O. Anderson, P. K. Sim, "Estimation of structured covariances with applications to array beamforming", Circuits, Systems and Signal Processing, Vol. 6, No. 4, 1987, pp. 421-447.
- [7] Y. I. Abramovich, D. A. Gray, A. Y. Gorokhov, N. K. Spencer, "Positive-definite Toeplitz completion in DOA estimation for non-uniform linear antenna arrays - part I : fully augmentable arrays", IEEE Transactions on Signal Processing, Vol. 46, No. 9, September 1998, pp. 2458-2471.
- [8] J. P. Burg, D. G. Luenberger, D. L. Wenger, "Estimation of structured covariance matrices", Proceedings of the IEEE, Vol. 70, No. 9, September 1982, pp. 963-974.
- [9] L. Vandenberghe, S. Boyd, "Semidefinite Programming", SIAM Review, Vol. 38, No. 1, March 1996, pp. 49-95.

# A HIERARCHICAL ALGORITHM FOR NEARFIELD ACOUSTIC IMAGING

Michael Peake<sup>†</sup>   Mehmet Karan<sup>†</sup>   Doug Gray<sup>†,‡</sup>

<sup>†</sup> Cooperative Research Centre for Sensor Signal and Information Processing (CSSIP)  
1 Warrendi Rd. Mawson Lakes, SA 5095, Australia

<sup>‡</sup> Department of Electrical and Electronic Engineering, University of Adelaide, Australia  
mpeake@cssip.edu.au   mehmet@cssip.edu.au   dgray@cssip.edu.au

## ABSTRACT

This paper presents a computationally efficient hierarchical nearfield imaging algorithm using a delay and sum beamformer with a random array via pulse-echo techniques.

In the conventional imaging of nearfield sources, signals from multiple sensors are delayed and summed to focus at points in the imaging volume. However, the computational burden of implementing the conventional algorithm for arrays with a large number of receivers, due to the distance calculations, requires new suboptimal algorithms. The algorithm described in this paper reduced computational time balanced against a reasonable memory requirements. The key principle of this algorithm is to implement the delay and sum by grouping receivers into various subarrays in a number of stages. Mirroring the hierarchy of subarrays is a decomposition of imaging volume beginning with large voxels and progressively reducing the voxel size. This algorithm introduces two main changes to the conventional algorithm which are (i) subsampling of the received data, which is then compensated by phase interpolations, (ii) a hierarchy of subarrays and subvolumes to combine the data from all the receivers.

Comparison of the conventional and the hierarchical algorithms are done in terms of their point spread functions.

## 1. INTRODUCTION

In this paper, we consider the problem of nearfield imaging using an acoustic transmitter with a bandwidth of several MHz and a receiving array of several thousand randomly positioned sensors. Acoustic imaging has found applications in diverse areas such as medical diagnosis, oceanic search, non-destructive evaluation and exploration seismology [1]. When the

object is assumed to be in the farfield of an array, the received waveform from a single point reflector can be assumed to be planar. Farfield approximation usually starts around at a range  $r = L^2/\lambda$  as described in [2] where  $r$  is the distance from the array,  $L$  is the diameter of the array and  $\lambda$  is the wavelength of the transmitted signal. However, when the object to be imaged is within this distance, new techniques such as in [3] need to be developed.

In underwater acoustic imaging, a common technique is to use a delay and sum beamformer. This technique obtains the image of an object from the delayed and summed backscattered echoes of a pulse modulated signal emitted by a transmitter. A carrier frequency in the region of 3-5 MHz is a compromise between low angular resolution at the lower frequency and high absorption at higher frequencies. As discussed in [4], use of a linear FM signal or "chirp" allows lower power transmission for a given range resolution and a bandwidth of several MHz allows millimetre resolution to be achieved. To obtain good cross-range imaging, a large receiver aperture array is required - typically of the order of 1 m<sup>2</sup>. The number of sensors required to densely fill such an aperture is prohibitive and thus a sparse array must be considered. The sensor locations are chosen randomly to reduce grating lobes that may otherwise occur (see [5, 6] and the references therein).

The computational burden of conventional time delay and sum imageformers using outputs of all the receivers is high [7]. Two features conspire against us. Firstly, the sparseness of the array means that subarrays have extremely directional beam-patterns and high sidelobes. If we combine receivers into a subarray by delaying and summing their data streams, that sum will only generate a good quality image of a small patch. Thus, this process must be repeated for each small patch to cover the full image. The second feature that makes the task difficult is that the object to be imaged is in the nearfield. Approximations that can be used in the farfield do not all apply here. For ex-

THIS WORK WAS SPONSORED BY THOMPSON MARCONI SONAR PTY LTD, AUSTRALIA.

ample, when calculating the distance from a receiver to many voxels in the image, either a time-consuming formula or a vast lookup table is required.

In acoustic imaging [4], an object is usually modelled in two ways: the object region is composed of a number of either point reflectors or several homogeneous media. In this paper, we employ the first assumption. Therefore the image of the object is assumed to be a superposition of point spread functions of the point reflectors in the object model. Hence, point spread functions are used as quality measures for the acoustic imaging algorithms in this paper.

In this contribution, we propose three novel techniques to address this nearfield acoustic imaging problem.

First, a piecewise polynomial approximation is used to quickly generate the distance calculations on the fly.

Secondly, the nature of the outputs of the matched filter used in the ranging processing is exploited to allow the time delay and sum imageforming to be implemented using coarse time delays followed by a fine time delay adjustment using phase multiplications.

Finally, and the main contribution presented, the imageforming is carried out in a hierarchical manner by grouping receivers into various subarrays in a number of stages. Mirroring the hierarchy of subarrays is a decomposition of the imaging volume beginning with large voxels and progressively reducing the voxel size.

In the next two sections, we describe the conventional and hierarchical algorithms and address some implementation issues.

## 2. THE CONVENTIONAL NEARFIELD ACOUSTIC IMAGING ALGORITHM

In conventional nearfield imaging [4, 7], a wideband signal, typically a linear FM chirp, is transmitted and backscattered echoes are received by an array of sensors. The sensor outputs are sampled, Hilbert transformed to produce complex signals and "dechirped" by convolving with a complex replica of the transmitted signal. The typical range profile from a single receiver is shown in Figure 1. To image in cross-range at each voxel, the round-trip distance is first calculated from the transmitter via the voxel to each receiver. The corresponding sample is then chosen from the dechirped data received from each sensor. These samples are summed, and the absolute value of the sum is chosen to be the voxel intensity.

This conventional algorithm is not practical due to high computational resource requirements. Therefore, we have altered that algorithm to achieve a balance between the speed, memory size and the image quality.

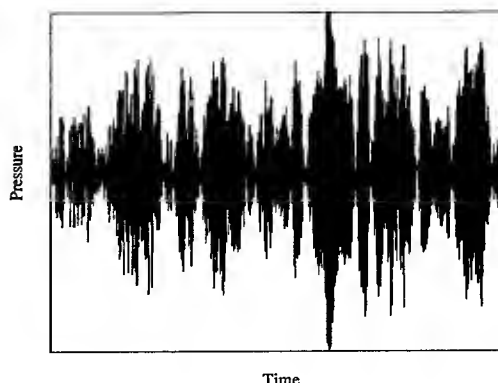


Figure 1: Range profile from a single receiver

The next section describes the new algorithm.

## 3. A HIERARCHICAL NEARFIELD ACOUSTIC IMAGING ALGORITHM

The hierarchical algorithm proposed in this paper divides the whole array into subarrays, and processes the data in a hierarchical manner. This algorithm is similar to the Quadtree Backprojection algorithm described in [8] for use with impulse radars and the fast beamforming algorithm in [9]. Depending on the number of algorithm stages chosen, the array is first divided into a number of subarrays of roughly equal area. These are divided into smaller arrays until we reach individual receivers. For example, if the algorithm is implemented such that it has three stages, then it divides the array into only small and large subarrays. If  $n_1$  large subarrays are used and each large subarray contains  $n_2$  points which are the centre points of small subarrays, then these  $N_2 = n_1 n_2$  points would form the centres of small subarrays. By varying the position and number of large and small subarrays, one can change the performance and the computational complexity of the hierarchical algorithm.

The hierarchical algorithm uses the key idea that a subvolume of many voxels will lie in the farfield of a small subarray, allowing shortcuts in the computation. Smaller subvolumes lie in the farfield of larger subarrays. Thus we deal with either fewer subvolumes and all the array, or all the voxels and a few large subarrays. By combining receivers into small subarrays and then into large subarrays, we can focus on first large image subvolumes and then on smaller ones and finally on voxels. This process is similar to the butterfly structure in FFT or the Butler matrix used in radar. However, in these cases where we are dealing with plane waves the butterfly decomposition is exact, whilst in

this approach it is approximate. Therefore, the proposed algorithm has a hierarchical structure where the concurrent contraction in the imaging subvolumes and an expansion in array subareas continue with an improvement in image resolution at each stage. This fact allows us to reduce the number of computations at each stage by a significant factor. A pseudo-code of the operations carried out in a stage of the algorithm can be seen in Table 1. In order to further reduce the algo-

```

for each large subarray
  load cubic coeffs for  $n_i$  small subarrays
  for each large subvolume
    for each of the  $n_i$  small subarrays
      load data stream, apply phase-shifts
      for each small subvolume
        Calculate distance
        Convert to (sample number, phase)
        Add data substream to new substream
      save new substreams
    zero new substreams for next large subvolume

```

Table 1: Pseudo-code of one stage of the hierarchical imaging algorithm

rithm requirements, both in computation and memory, the two techniques described in the following section can be used. These techniques may also be incorporated in the conventional algorithm.

#### 4. IMPLEMENTATION ISSUES

In many software and hardware implementations, excessive delays can be caused by pre-calculating delay values and transferring the required blocks of data and the delays in and out of cache memories. It is often more efficient to calculate time delays on the fly which requires fast algorithms. The following technique can be used to do so and it will be employed in the new hierarchical algorithm described in the next section.

The distance can be approximated well over part of the imaging volume by a cubic polynomial. We will use several polynomials  $P_i$ , each covering a part  $V_i$  of the imaging volume. The coefficients in each polynomial  $P_i$  are found by regression using several distances within  $V_i$ . This can be done off-line and only twenty coefficients per cubic polynomial need be stored.

To evaluate each cubic polynomial over many equally spaced voxels, we use the following method. We could evaluate a linear function  $f(n) = a + bn$  at integers  $n = 0, 1, 2, \dots$  simply by starting with  $a$  and continually adding  $b$  to the previous answer. This uses one addition to evaluate  $f(n)$  for each value of  $n$ . In the

same way, by nesting these additions three levels deep we can evaluate a cubic polynomial using just three adds for each value of  $n$ .

In addition to distance approximations, extra speed-up of acoustic imaging algorithms may be possible via subsampling and then interpolating received signals. In many applications, design considerations dictate a sampling rate that exceeds the transmitted signal bandwidth by a factor of 4-10. The impulse response of the matched filter output (ie. the range compressed chirp) typically looks like Figure 2. Time delaying a receiver

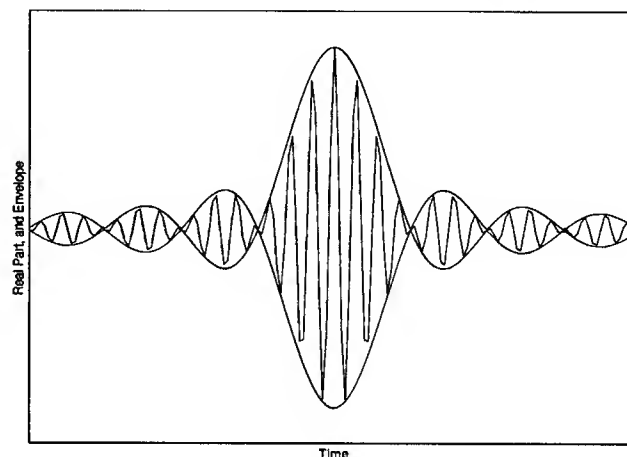


Figure 2: The output of a filter matched to a linear FM signal driven by the same signal.

output can be achieved by first selecting at a suitable subsampling factor the sample nearest to the peak of the modulation - this is termed coarse time delay. Provided we are within the 3 dB point of the peak, a fine time delay adjustment may be made through multiplication by the appropriate phase factor. This subsampling (typically 4-8 in practice) allows substantial savings in down-stream computations as it enables the hierarchical algorithm of the following section to be implemented.

The above techniques can be used also with the conventional algorithm. A combination of these ideas are used in the proposed algorithm to reduce the number of calculations to a small fraction of the number needed by the conventional acoustic imaging algorithm.

#### 5. SIMULATION STUDIES

Figure 3 gives the cross-sections of the point spread functions of the conventional and hierarchical algorithms in the  $(x, y)$  plane, while Figure 4 shows similar cross-sections in the  $(x, z)$  plane. In this example, an ar-

ray of around 2500 receivers pseudo-randomly spaced in a (50cm)x(50cm) aperture was used, implying an average separation of about  $10\lambda$ . A three-stage hierarchical algorithm was employed with 21 large subarrays and sixteen small subarrays in each large subarray. The decomposition of the image volume was 9, 3 and 1mm respectively at a range of one metre. The subarrays, as well as the receivers, formed irregular patterns. We have found that regular subarrays gave rise to high sidelobes, even if the receivers had no regularity at all.

Figure 3 shows a raised sidelobe floor in the region near the mainlobe. The value of the point spread function is much higher in the central subvolume than in the surrounding subvolumes. Also, the position of the point reflector relative to the coarse subvolumes affects the point spread function. Figures 3 and 4 show the result when the point reflector happens to be at the centre of a coarse subvolume, which is when the algorithm does best. Greater errors are introduced by the algorithm when the point reflector is the edge of a coarse subvolume. The cross-section in the range direction can be seen in Figure 4. It can be observed that most of the blurring is in the cross-range direction. The hierarchical algorithm causes sidelobes to blur more in the range direction, while as Figure 3 indicates, the mainlobe is blurred mostly in the cross-range direction.

## 6. CONCLUSION

In this paper, we have presented a hierarchical method for nearfield acoustic imaging using a large, pseudo-random array of sensors. This algorithm achieves a slightly degraded image with a reduced amount of memory in a shorter computation time. The algorithm applies several modifications to the conventional imaging algorithm. These modifications are as follows:

- The use of a hierarchy of subarrays allows a significant reduction in the number of calculations required. The amount of time saved depends on the resolution required compared with the density of receivers in the array. However, close attention must be paid to the arrangement of subarrays to limit degradations to the image.
- The data was down-sampled to reduce the number of operations required. The resulting image would be severely degraded unless interpolations were applied in the form of complex phase-shifts. We allowed only a quantized set of phase-shifts. These phase-shifts were found efficiently, and did not cause a great degradation in the image.
- The calculation of time-delays, which forms a large part of the conventional algorithm, was achieved

in the hierarchical one with polynomial approximations in a few adds each.

Moreover, the hierarchical algorithm allows minimal interaction with the hard disc, which is large enough to contain all the data but slow to fetch data from. This is as important as the reduction in number of calculations.

Point-spread functions were presented to compare the two algorithms. Raised sidelobe floors were seen in the hierarchical algorithm's point spread function. These can be reduced by a judicious choice of irregular subarrays.

## REFERENCES

- [1] H. Lee and G. Wade, eds., *Modern Acoustical Imaging*. IEEE Press, 1986.
- [2] A. Macovski, "Ultrasonic Imaging Using Arrays," *Proceedings of IEEE*, vol. 67, pp. 484-495, April 1979.
- [3] R. A. Kennedy, T. D. Abhayapala, and D. B. Ward, "Broadband Nearfield Beamforming using a Radial Beampattern Transformation," *IEEE Transactions on Signal Processing*, vol. 46, no. 8, 1998.
- [4] M. Soumekh, *Fourier Array Imaging*. Prentice-Hall, Inc., 1994.
- [5] V. Murino, A. Trucco, and A. Tesei, "Beam pattern formulation and analysis for wide-band beamforming systems using sparse arrays," *Signal Processing*, no. 56, pp. 177-183, 1997.
- [6] S. Holm, B. Elgetun, and G. Dahl, "Properties of the Beampattern of Weight- and Layout-Optimized Sparse Arrays," *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*, vol. 44, no. 5, pp. 983-991, 1997.
- [7] D. G. Blair and I. S. F. Jones, "Underwater acoustic imaging: Rapid signal processing," Tech. Rep. DSTO-TN-0098, Maritime Operations Division, Aeronautical and Maritime Research Laboratory, DSTO, January 1998.
- [8] J. H. McClellan, S.-M. Oh, and M. C. Cobb, "Quadtree Focusing for UWB SAR," in *Defense Applications of Signal Processing Proceedings of the 1999 Workshop*, (LaSalle, Illinois, USA), 1999.
- [9] K. Houston, "A Fast Beamforming Algorithm," in *Proceedings of OCEANS 94. Oceans Engineering for Today's Technology and Tomorrow's Preservation*, vol. 1, pp. I/211-16, IEEE, 1994.



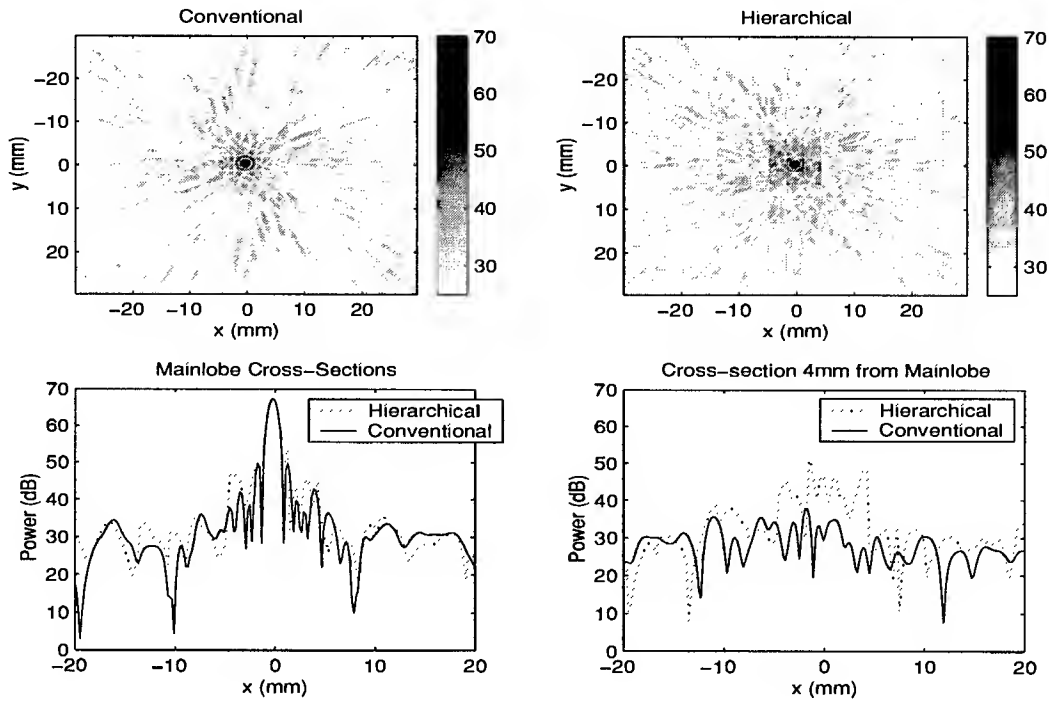


Figure 3: Conventional and hierarchical algorithm point spread functions in the  $(x, y)$  plane. Intensities shown in dB scale. The lower graphs show the cross-sections of the point spread functions of hierarchical and conventional algorithms at  $y = 0$  and  $y = 4$  mm.

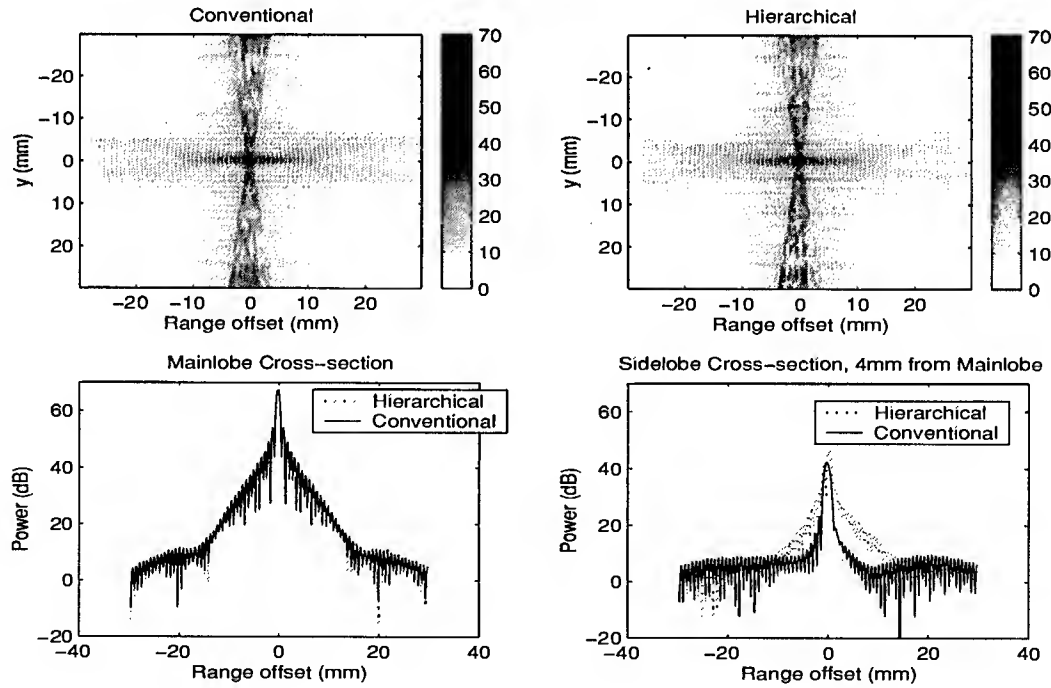


Figure 4: Conventional and hierarchical algorithm point spread functions in the  $(x, z)$  plane. Intensities shown in dB scale. The lower graphs show the cross-sections of the point spread functions of hierarchical and conventional algorithms at  $y = 0$  and  $y = 4$  mm.



# AN INTRODUCTION TO SYNTHETIC APERTURE SONAR

David Marx, Matt Nelson, Enson Chang, Walt Gillespie, Angela Putney, Kieffer Warman

Dynamics Technology, Inc.  
21311 Hawthorne Blvd., Suite 300  
Torrance, California 90503

## ABSTRACT

DTI has demonstrated that Synthetic Aperture Sonar (SAS) can provide orders of magnitude improvement in cross-range resolution when compared to conventional sonar beamforming. Like Synthetic Aperture Radar (SAR), SAS processing requires coherence over multiple measurements, but has long been impractical due to the nature of the ocean environment. We have extended SAR processing ideas to accommodate the issues specific to the underwater environment, and have successfully synthesized apertures extending many thousands of wavelengths. We will present an overview of the theory of SAS processing, how it differs from SAR, and will show experimental results from SAS processing of sonar data.

## 1. INTRODUCTION

Synthetic Aperture Sonar (SAS) is an underwater adaptation of the techniques used in Synthetic Aperture Radar (SAR). SAS is the coherent integration of data from a number of transmissions as the sonar moves along its track. The resolution limit for the focused image in the cross-range direction is one half the lateral size of the receiver element, at all ranges. High area coverage rate is achieved by use of a multiple element receive array.

Not only does SAS allow for the generation of sonar imagery with range-independent resolution, but the resolution is also independent of frequency, up to normal diffraction limits. Synthetic aperture processing is scalable, and it has been applied in both mine hunting and anti-submarine warfare (ASW) scenarios.

When adapting SAR algorithms to SAS systems, several important differences between the two technologies come to light. Table 1 lists some of them. While wavelengths are similar for the two systems, SAS systems typically use radiating elements smaller, in proportion to the wavelength, than SAR systems. As a result, radiation beampatterns for SAS systems have a greater angular extent than those for SAR systems. The effect on synthetic aperture processing is that range migration (the shifting of a target's return through range resolution cells as the platform flies past) is quite significant and pronounced for SAS systems.

Because of the long propagation delays and the desire to minimize the synthetic aperture time, SAS systems generally use an array of receiver elements. Aperture sampling requirements, along with the slow propagation, imply an interdependence between array length ( $L$ ), ping repetition interval (PRI), maximum platform velocity ( $V_{SAS}$ ) such that,

$$V_{SAS} \leq \frac{L}{2PRI}$$

Perhaps the most significant difference between SAS and SAR systems is the medium coherence. Along with medium fluctuations, platform stability can also affect the signal phase history. Since synthetic aperture times for SAS are an order of magnitude greater than those for SAR, the phase corruption due to these effects can be much more severe for SAS.

	SAR	SAS
$c$ [m/s]	$3 \times 10^8$	$1.5 \times 10^3$
wavelength [m]	0.01 – 0.3 (typ.)	0.01 – 1.5
SA time	few seconds	minutes – hours
SA length	few km (SIR)	30 m – 5 km
PRI	0.5 – 1 ms (SIR)	0.1 s – 1 min.
range stand-off	> 225 km (SIR)	few meters – km
medium coherence	Days	minutes

Table 1. Comparison of SAR and SAS key parameters. Some of the SAR parameters are derived from the Shuttle Imaging Radar (SIR).

In the next section we describe the flow of a SAS processor, and in Section 3, we present results of three different SAS systems. The DARPA SAS system and the CSS system are intended for mine countermeasures and use rigid tow bodies. The SWAC data was collected using a long flexible towed array at low frequencies.

## 2. SAS PROCESSING

Synthetic aperture sonar produces high-resolution imagery by coherently combining data from multiple sonar pings. To produce high-quality SAS results, these data must be referenced to a straight platform track with sub-wavelength accuracy. The motion of even well-behaved platforms, and phase distortions from the complicated underwater environment conspire to make this a challenging requirement, and much of our SAS processor is devoted to overcoming platform motion and phase errors. In general, our synthetic aperture processing can be divided into three, usually distinct, operations. In the first step, the data are adjusted to compensate for initial estimates of platform motion, using inputs from both the hardware mocomp suite, if one exists, and from information extracted from the sonar data itself. The

second component is the image formation algorithm that performs azimuthal compression on the motion-compensated raw data and generates an image from the raw data. The third operation involves correction of phase errors due to *uncompensated* platform motion and medium instabilities, which is usually accomplished by application of some form of autofocus algorithm

## 2.1 Motion Compensation

When available, platform motion estimates derived from specialized motion compensation hardware are applied to the sonar data to remove many of the effects of platform motion. Unfortunately, such estimates are often not available, due to hardware failures or simply the absence of these costly instruments. To process data with this limitation, we estimate phase error through the analysis of the of the sonar returns only. One technique we employ to estimate towbody motion from sonar data is a Prominent Point Processor [1][2], which relies upon the existence of a dominant scatterer in the scene. After manually selecting the prominent point and locating its point of closest approach, its echo history is compared to the theoretical form it would have were there no platform motion. At each along-track location, the discrepancy in range and phase is calculated and is interpreted as a measurement of towbody displacement. This set of motion estimates is then applied to the entire data set, restoring the sonar returns to their theoretical zero-motion locations.

Though robust, this technique requires the existence of strong, point-like scatters throughout the area being imaged. Moreover, the returns from those prominent points must be distinguishable from other echoes throughout the whole integration length that it takes to develop a synthetic aperture. An alternative method for estimating platform motion from sonar scene content, known as Redundant Phase Center processing, has also been tested. RPC is derived from the SAR technique of along-track interferometry. RPC estimates towbody motion by performing range-wise correlations on data from overlapping segments of the hydrophone array on successive pings of the sonar. Our RPC implementation for SAS requires neither a prominent point nor critical decisions and input from the analyst. Since the received signal varies as  $e^{i\omega t}$ , where  $\omega$  is the carrier frequency, the *complex* correlation between the two repeated looks is sensitive to phase errors on the order of a fraction of a cycle.

Although either of these algorithms could fail if the sonar data is of poor quality, they do have advantages over hardware-based motion measurement approaches. By its nature, motion measurement hardware can provide information about platform motion only. Unfortunately, there are many other sources of phase corruption in the underwater environment, such as medium instabilities, multipath propagation, and medium inhomogeneities. Phase errors derived from analysis of sonar echoes include the effects of all of these sources of phase corruption, and can therefore correct all phase errors simultaneously.

## 1.2 RMA Image Formation

Several SAR image formation algorithms exist in the literature. Examples are direct matched filtering, chirp scaling algorithm, and the range migration algorithm (RMA) [2][3]. Because of its speed and full 2-dimensional treatment of the problem, we have chosen RMA the image formation stage of our SAS processor, and the images shown in the next section were all focused using RMA. Developed for work in the geophysics community, RMA implements an exact solution of the synthetic aperture problem using fast Fourier transforms to efficiently apply the matched filter in the wavenumber-frequency domain. With this technique, both speed and theoretical performance are achieved.

Since the RMA formalism explicitly treats only single-element systems, where the transmitter and receiver use the same antenna. However, a typical SAS system uses a single transmitter with an array of receivers, and in addition, the transmitter may be physically separated from the receiver array by a significant distance. Therefore, pre-processing, in the form of a phase correction, is required to correct for the physical separation between each receiver and the transmitter. After this correction, the receiver data is equivalent to samples along the synthetic aperture corresponding to a single transmitter/receiver antenna.

## 1.3 Phase Gradient Autofocus

To remove residual uncompensated motion from the focused images, we employ the phase gradient autofocus (PGA) algorithm [4][5]. The PGA algorithm selects candidate point-like targets in the synthetic aperture image, estimates the residual phase error at those points, and combines them into an optimal estimate of the phase error. This phase error is removed from the image and the process iterated until convergence is obtained.

To estimate the phase error, only strong scatterers are used. Once detected, they are windowed and shifted to occupy same location in the synthetic aperture (doppler history). The phase gradient is estimated by,

$$\dot{\phi}(u) = \frac{\text{Im}\{G^*(u)\dot{G}(u)\}}{|G(u)|^2},$$

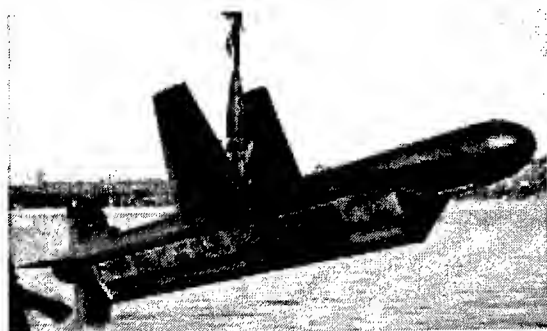
where  $G(u)$  is the signal along the synthetic aperture,  $u$ . Other phase gradient estimators exist, but the above estimator is the linear unbiased minimum variance estimator. Integrating the phase gradient is then an estimate of the phase error. After correcting the data for the estimated phase error, the algorithm can be iterated.

## 3. EXPERIMENTAL RESULTS

### 3.1 DARPA SAS System

The DARPA SAS system (Figure 1), built and operated by Raytheon, is a heavy-tow body with a one-sided, keel-mounted hydrophone array consisting of 32 elements. The contiguous array elements each have a width of 10.9 cm, and were originally configured with half of the elements forming a 16-element array, allowing for a final cross-range resolution of 5.5 cm. Later, they

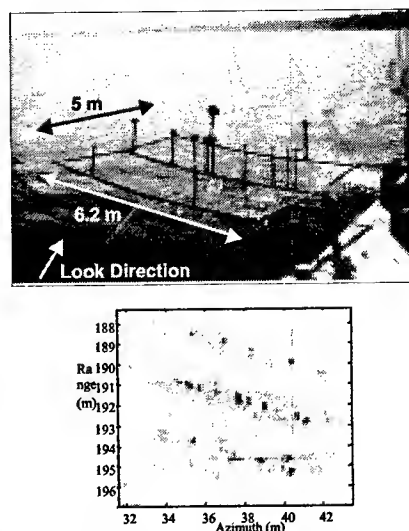
were wired in pairs to produce a 16-element array of 21.8 cm elements, supporting a resolution of just under 11 cm. The projector array is located on an adjustable "wing" approximately 0.8 m above the receiving array. This design makes it possible to transmit a narrow vertical beam to reduce the effects of multipath and concentrate transmitted power on long-range targets. The projector operates on a center frequency of 50 kHz, and can transmit in either of two modes: a coherent burst consisting of 6 cycles of the carrier frequency, or a linear FM (LFM) chirp of programmable duration and bandwidth. With a bandwidth of 10 kHz, the LFM mode yields a theoretical range resolution of 7.5 cm; the 6-cycle tone achieves 9 cm.



**Figure 1.** Raytheon-DARPA SAS towbody. Keel-mounted hydrophone array is 3.2 m long. Projector array is located in adjustable wing and provides narrow vertical beam for long-range operation.

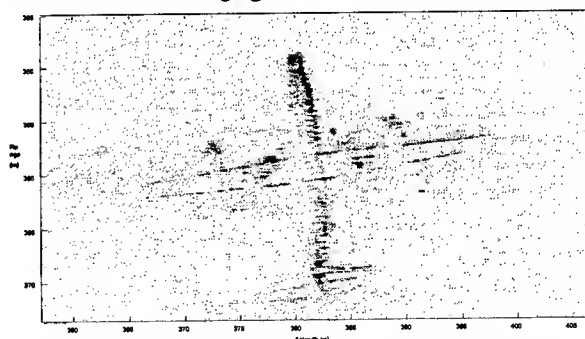
A PVC frame supporting a variety of known scientific targets was deployed during several sea tests (Figure 2). Once installed, the rectangular frame lies flat on the sea bottom, and has a series of posts protruding upward to support targets. Posts running down the center of the frame supported two groups of five air-filled steel spheres: one group consisted of 10-cm diameter spheres, the other 20-cm. In each group, the targets were separated by 1, 2, 4, and 8 diameters. The spheres should behave as ideal, but low cross-section, point scatterers and thus test the resolution of the SAS system. Two posts along the farthest edge of the frame supported triplane corner reflectors (30- and 60-cm diameter), while two posts on the closest edge were empty.

Figure 2 shows the final SAS image obtained from data collected at a range of about 190 m in Lake Washington, with the array configured with 10.9 cm receiver elements. This image was generated using RPC, RMA, and PGA, as described in the last section. Visible in the center of the SAS image are the spheres (20 cm on the right, 10 cm on the left), as well as a corner of the support frame at the extreme right. The artifacts appearing as double-images in range of the spheres is consistent with what would be expected from a single multipath bounce off the bottom of the lake. At the far edge (bottom), the two strong corner reflectors are evident, as is another corner of the frame. The empty posts along the near edge (top) are also visible. The support frame itself is visible as a faint line connecting the target images. The 3-dB down-point resolution achieved in this image, evaluated by taking a series of intensity cuts through the images of the 10- and 20-cm spheres, varies slightly within the image, with the best cross-range resolution being approximately 7 cm, compared to 5 cm theoretical.



**Figure 2.** Target frame (top) with 10- and 20-cm spherical targets centrally mounted. Conventional beamformed image (lower left) shows much less detail than SAS image (lower right).

A popular sonar target in Lake Washington is a sunken PB4Y-2 navy patrol aircraft resting in 50 m of water. A SAS image of this target obtained from a range of 350 m and processed with prominent point motion estimation is shown in Figure 3. The most striking feature compared with other published sonar images of this target is that the aircraft has a ghostly appearance, as if its skin has been removed. Subsequent analysis has determined that the thin aluminum skin, immersed in water on both sides, is nearly transparent to the 50 kHz sonar signals used by this sonar, with less than 2 dB of transmission loss through the skin decreasing to virtually no loss at grazing incidence. Comparisons to published plans of the PB4Y-2 confirm that the periodic features seen along the fuselage in the SAS image are interior structural elements of the airplane. The wing and tail spars are visible, as are some of the ribs that support the aft fuselage. The locations of highlights in the tail portion of the image agree with the blueprints showing the locations of frames in the PB4Y-2 to within 3%, thus supporting the conclusion that the 50 kHz sonar is imaging the interior of the water-filled target.



**Figure 3.** PB4Y-2 airplane from a range of approximately 350 m. The SAS image shows details of the planes internal structure, resolved to approximately 15 cm.

A second run past the airplane collected data from a range of 980 m. A SAS image produced from this data is shown in Figure 4. Although the long-range image is not illuminated as well as the airplane seen from closer range, the resolution is approximately the same in the two images. Sound velocity profiles in the lake indicated that medium was downward-refracting, and ray tracing codes indicated that there were no direct acoustic paths past ranges of about 500 m, where the last ray hit the bottom. The airplane was therefore imaged with sound that had suffered two bottom bounces. In fact, the airplane is situated such that the illuminating ray apparently came from a side-lobe of our narrow-vertical-beam-width transmitter and even then it almost missed the aircraft.

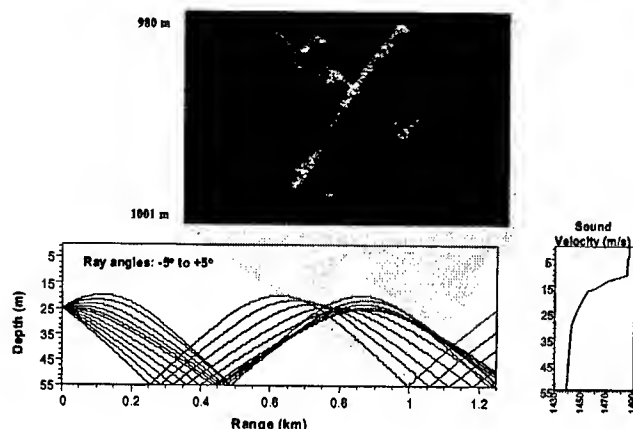


Figure 4. SAS image of PB4Y-2 airplane at 980 m (above). Sound velocity profile and ray trace results for Lake Washington.

### 3.2 CSS High and Low Frequency SAS Results

Coastal Systems Station (CSS) in Panama City, Florida collected SAS data using an array built by Northrop-Gruman. This array contained 14 high frequency (180 kHz) elements and 7 low frequency (20 kHz) elements. The length of the high frequency elements was 5.08 cm, giving a theoretical cross-range resolution of 2.54 cm, and the high frequency elements were 3.81 cm. However, the cross-range resolution for the low frequency is limited by the wavelength (7.5 cm) rather than the element length. For these experiments, a series of objects, such as a large cylinder, a ladder, and a truncated cone, were placed on a sandy bottom. Figure 5 shows the focused image for the high frequency data, and Figure 6 is the resulting image for the low frequency. In the case of the high frequency data, the image was formed using RPC and RMA, but an autofocus step was not required. The low frequency image was formed using all three steps.

Some interesting differences between the two images indicate how target response is frequency dependent. For example, the cylinder object (upper right corner) in the high frequency image is bright along its whole length, and a distinct shadow is visible. However, the only its ends are bright in the low frequency image, and no shadow is present. Similarly, the long shadow behind the truncated cone (just below the cylinder) also disappears for the low frequency image.

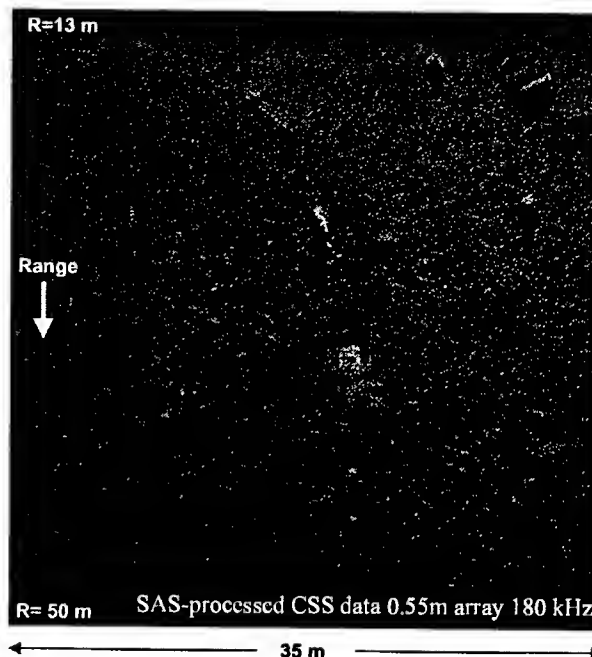


Figure 5. High frequency (180 kHz) SAS image. A cylinder in the upper right gives a clear shadow.

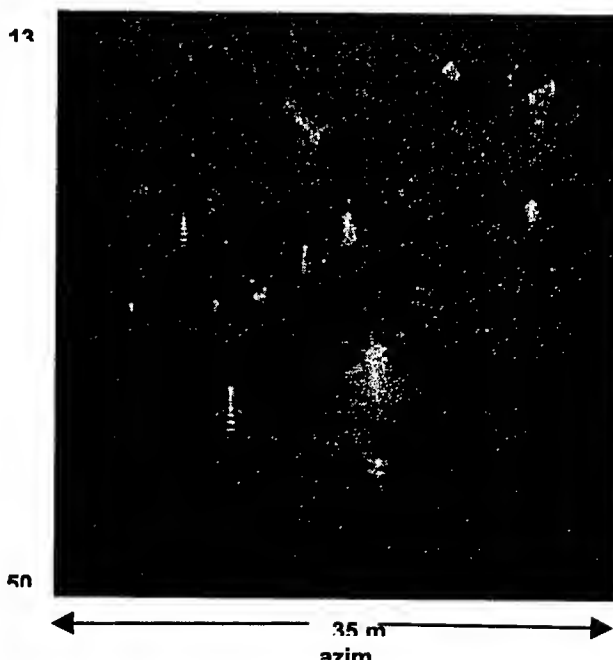


Figure 6. Low frequency (20 kHz) SAS image. Objects and background reverberation have different image characteristics at 20 kHz than at 180 kHz.

### 3.3 ONR SAS Results

The low-frequency sonar system used in the SWAC tests consists of two separately towed components, a 900 m neutrally-buoyant flexible line, the central 256 m of which are populated by 256 hydrophones, and a heavy-towed transmitter vehicle with a 10 m vertical projector array. The hydrophones are not uniformly spaced along the array, so we resample the data onto a uniform grid prior to our SAS processing. The system operates at a 600 Hz center frequency, and is capable of ensonifying targets at very long ranges. Two-vehicle systems such as this are less than ideal for SAS, because the transmitter and receiver positions are not constrained relative to one another as they are in a single-vehicle sonar system. As a result, the two-vehicle configuration adds another degree of freedom to the motion compensation problem. An additional issue with this data set had to do with speed of advance. To ensure proper spatial sampling of the synthetic aperture, the array can advance no farther than one-half of its length on every ping. During the SWAC trials, the speed of advance exceeded this SAS speed limit by approximately 30%. In spite of these issues, we were able to produce focused SAS images from these data, albeit with somewhat elevated side-lobe levels due to the tow speed excess.

The SWAC data set contained a strong return at a range well beyond the test area. Later investigation revealed the presence of an oil rig in this approximate location, which we now believe to be the source of these long-range returns. Figure 7 shows our SAS image of this target. The SAS image was generated after prominent point corrections from a nearby return were applied to the data; autofocus was also applied to the SAS result. Visible in this image are distinct highlights (probably echoes from the oil rig's support structure) resolved to approximately 5 m. In contrast, a conventional beamformed image using the real aperture of the flexible towed array would have a cross-range resolution of about 250 m. Because the sonar was towed at a speed higher than the SAS limit, cross-range grating sidelobes are also clearly visible in this image.

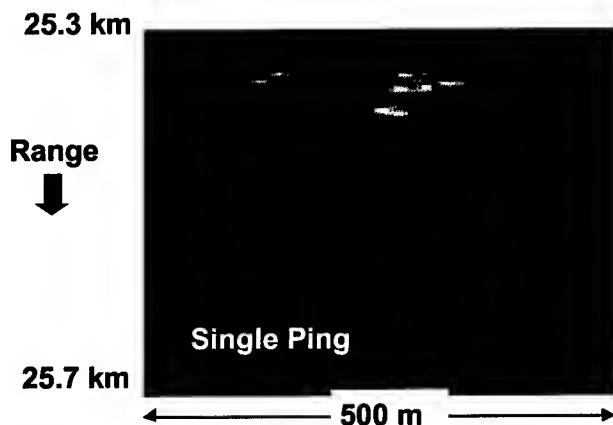


Figure 7. SAS image generated from the same data (right) shows 5 m resolution. Also visible in the image are azimuthal sidelobes, a consequence of the sonar advancing faster than the SAS speed limit.

### 4. CONCLUSIONS

We have demonstrated that synthetic aperture imaging is capable of providing order-of-magnitude improvements in cross-range resolution over conventional sonar beamforming techniques. SAS techniques are widely applicable to sonar systems of widely varying characteristics, and appear to be robust in the face of multipath acoustic environment. Table 2 summarizes the SAS resolution results for the three systems presented in this paper.

	wave-length	resolution limit	processor resolution	measured resolution
DARPA	3 cm	5.0 cm	7.0 cm	7.0 cm
	3 cm	10.0 cm	15.0 cm	17 cm
CSS high	0.83 cm	2.54 cm	4.3 cm	5.3 cm
CSS low	7.5 cm	7.62 cm	11.4 cm	15 cm
ONR	2.5 m	5.0 m	5.0 m	~11 m

Table 2. Summary of resolution results for several experimental SAS systems.

### 5. REFERENCES

- [1] Cohen D. and Nelson M. "ARPA SAS Phase 1 Final Report". Dynamics Technology, Inc. internal report DTW-9322-96012, September 1996.
- [2] Carrara W., Goodman R., and Majewski R., *Spotlight Synthetic Aperture Radar*. Boston, MA: Artech House, 1995.
- [3] Cafforio C., Prati C., and Rocca E., "SAR Data Focusing using Seismic Migration Techniques". *IEEE Trans. on Aerospace and Electronic Systems* 27, pp. 194-206, 1991.
- [4] Eichel P., Ghiglia D., and Jakowatz C., "Speckle Processing Method for Synthetic Aperture Radar Phase Correction". *Optics Letters* 14, pp. 1-3, 1989.
- [5] Wahl D.E., Eichel P.H., Ghiglia D.C., and Jakowatz C.V., "Phase Gradient Autofocus—A Robust Tool for High Resolution SAR Phase Correction". *IEEE Transactions on Aerospace and Electronic Systems*, Vol. 30, No. 3, pp. 827—835, 1994.

# CLASSIFICATION OF ACOUSTIC AND SEISMIC DATA USING NONLINEAR DYNAMICAL SIGNAL MODELS

Ron K. Lennartsson

Defence Research Establishment  
SE 172 90 Stockholm, Sweden.  
e-mail: ron@sto.foa.se

Áron Péntek and James B. Kadtke

Marine Physical Laboratory  
Scripps Institution of Oceanography  
University of California, San Diego, USA.  
e-mail: {apentek,jkadtke}@ucsd.edu

## ABSTRACT

One of the most important applications of nonlinear dynamics is the estimation of empirical dynamical models from data, in order to explain time series derived from physical processes. Such derived models can then be used for a variety of data processing applications, in particular for detection and classification problems. Typically, the parameters of such dynamical models are estimated directly from the time series by minimizing a cost function with least squares. In this paper we discuss the theory and applications of an alternate approach for estimation of such nonlinear dynamical models and the use of these models for detection and classification of seismic and acoustic data. We apply these ideas to real data derived from seismic station recordings in the region of the Panama Canal. Finally we compare our results with that previously achieved by the method of Master-event correlations, and find improved performance. This indicates that a dynamical model approach incorporates additional signal information in this example.

## 1. INTRODUCTION

In recent years several attempts have been made to incorporate advances in nonlinear dynamical systems theory into robust signal processing tools for analyzing complicated time series. In this paper we develop a method for the estimation of delay differential equation signal models, motivated by the Yule-Walker equations of autoregressive modeling theory [1]. Estimating the model in this way involves estimation of both higher order statistical moments and dynamical moments. The dynamical moments are also of higher order, but involve the derivative of the signal [2]. We demonstrate that these models can simultaneously incorporate standard linear and nonlinear signal measures. In addition they also express information directly related to low-dimensional deterministic signal evolution. Our method applies to very general classes of dynamical models which may incorporate vector data streams of several physical observables, multiple time-delayed variables, and even explicitly non-stationary models. Here, however, for clarity and because of space limitations, we will summarize the basic technique on a sub-class of models consisting of delay-differential equations (DDEs) in a single scalar variable. DDEs are known to succinctly describe a wide

variety of physical processes, and their estimation has been of considerable interest recently [3][4].

Using this theory we outline the design for practical detection and classification algorithms for acoustic and seismic data analysis applications. We define a feature space from the model coefficients, and implement both Mahalanobis decision criteria and a neural network algorithm to generate signal class hypothesis testing. We discuss the scaling properties of this detector and compare it to a standard energy detector. We show that the dynamical detector scales with the sampling rate as a matched filter detector even though no exact signal template is used. Finally, we discuss the utility of the higher-order dynamical moments as classification features and compare our results to other classification algorithms which use higher-order moments for classification.

## 2. ESTIMATION OF NONLINEAR DYNAMICAL MODELS

As a general description, we first assume that we observe a continuous scalar data stream  $x(t)$  generated by measurement of some accessible observable of a physical process. We hypothesize that the process evolution itself can be approximated by a deterministic, relatively low-dimensional dynamics, but can include purely stochastic elements (i.e. noise) as well. We will also utilize up to  $D$  time-delayed copies of  $x(t)$ , written  $x(t - d\tau)$  with  $1 \leq d \leq D$ . Hence our general model form is

$$\dot{x}(t) = F[x(t), x(t - \tau), \dots, x(t - D\tau)]. \quad (1)$$

The function  $F$  is often expanded in terms of some basis functions. For our analysis we will restrict our attention to two-delay second order models of type

$$\dot{x} = a_1 x_{\tau_1} + a_2 x_{\tau_2} + a_3 x_{\tau_1} x_{\tau_2} \quad (2)$$

where we introduced the shorthand notations  $\dot{x} \equiv \dot{x}(t)$  and  $x_{\tau} \equiv x(t - \tau)$ . This model has been used successfully to model and detect quadratic phase couplings [1].

Here we diverge from an exact modeling approach which is often employed in nonlinear dynamics theory. For modeling purposes determination of a correct functional form  $F$  is necessary to recover exact dynamical information about the original system, which is typically problematic. However

for classification purposes, we postulate that it is only necessary to incorporate sufficient model dimensionality and nonlinearity to distinguish the required signal classes, regardless of the exact form of the original dynamical generators. Typically, we find that dimension  $D$  and the order of nonlinearity can be small (e.g. 2 or 3), since signal power is usually distributed mostly in the lowest orders of nonlinearity and dimension (in analogy with spectral expansions). More importantly, we find it crucial to define a standard model form for the dynamical model (i.e. a fixed "dynamical filter") in a particular application, otherwise comparison between different signal classes becomes impossible. The unknown model coefficients  $a_1$ ,  $a_2$ , and  $a_3$  are estimated for each data window, and will comprise our classification feature space. This estimation must be numerically robust and hopefully explicitly preserve some of the nonlinear correlations possibly present in the original signal.

Next we present a method which can accomplish both of these goals, and make explicit connection to higher-order spectral theory. Briefly we multiply Eq. (2) by each basis term  $x_{\tau_1}$ ,  $x_{\tau_2}$ , and  $x_{\tau_1}x_{\tau_2}$ , and average over an observation window of length  $T$ ; the model coefficients are then computed by solving the following linear equation:

$$\bar{R} * \bar{A} = \bar{B} \quad (3)$$

where

$$\bar{R} = \begin{pmatrix} \langle x^2 \rangle & \langle x_{\tau_1} x_{\tau_2} \rangle & \langle x_{\tau_1}^2 x_{\tau_2} \rangle \\ \langle x_{\tau_1} x_{\tau_2} \rangle & \langle x^2 \rangle & \langle x_{\tau_1} x_{\tau_2}^2 \rangle \\ \langle x_{\tau_1}^2 x_{\tau_2} \rangle & \langle x_{\tau_1} x_{\tau_2}^2 \rangle & \langle x_{\tau_1}^2 x_{\tau_2}^2 \rangle \end{pmatrix}$$

$$\bar{A} = \begin{pmatrix} a_1 \\ a_2 \\ a_3 \end{pmatrix} \quad \bar{B} = \begin{pmatrix} \langle \dot{x} x_{\tau_1} \rangle \\ \langle \dot{x} x_{\tau_2} \rangle \\ \langle \dot{x} x_{\tau_1} x_{\tau_2} \rangle \end{pmatrix}.$$

Where  $\langle \star \rangle$  stands for the expectation value. The linear system (3) are the normal equations for Eq. (2) similar to that of the Yule-Walker type equations in parametric modeling theory [9] where the model coefficients are expressed by the correlation matrix. Note that the correlation involving the signal derivative can be calculated from the derivative of the correlation function, i.e.

$$\begin{aligned} \langle \dot{x} x_{\tau_1} \rangle &= \frac{d}{d\tau_1} \langle x x_{\tau_1} \rangle, \\ \text{and} \\ \langle \dot{x} x_{\tau_1} x_{\tau_2} \rangle &= \frac{d}{d\tau_1} \langle x x_{\tau_1} x_{\tau_2} \rangle + \frac{d}{d\tau_2} \langle x x_{\tau_1} x_{\tau_2} \rangle. \end{aligned} \quad (4)$$

These formulas are valid in the long window limit for a bounded stationary signal  $x(t)$ . The main practical advantage of using Eq. (3) instead of solving Eq. (2) in a least squares sense is that we can avoid computing the signal derivatives, which is the main difficulty for noisy signals. The expectation values on the left hand side of

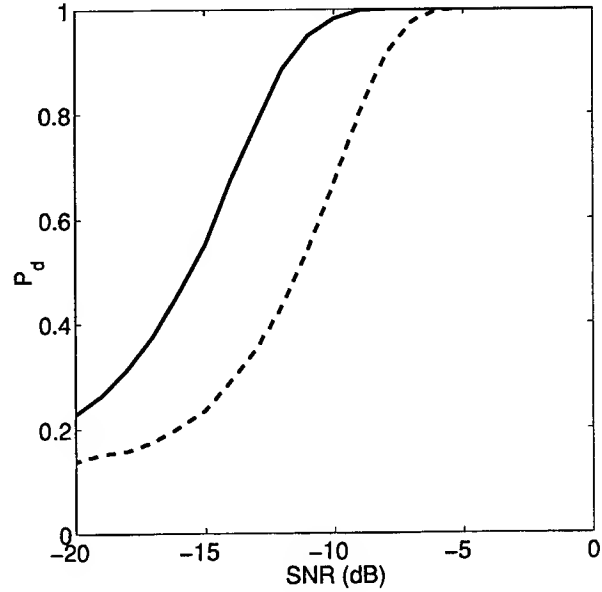


Figure 1: Numerically estimated detection probabilities  $P_d$  vs. signal to noise ratio (SNR) for the harmonic signal for the dynamical detector (solid line) and energy detector (dashed line).

Eq. (3) can be expressed as standard higher-order data moment functions [5]. For example,  $\langle x_{\tau_1} x_{\tau_2}^2 \rangle = m_{xxx}(\tau_2 - \tau_1, \tau_2 - \tau_1)$  where the 3rd order moment function is defined as  $m_{xxx}(\tau_1, \tau_2) = \langle x(t)x(t-\tau_1)x(t-\tau_2) \rangle$  and describes bi-correlations. We also note that the dynamical moments involving  $\dot{x}$  arise exactly because of the dynamical representation and express information not utilized in standard higher order methods.

We have applied the above classification methods to a variety of real world data sets, including stationary and transient sonar data [1] and dolphin echo-location data [6]. In this paper we apply these ideas to the automated classification of real-world seismic data derived from seismic station recordings in the region of the Panama Canal [7][8].

### 3. SIGNAL DETECTION

Let us consider the simple example of a harmonic signal  $x(t) = \sin(\omega t)$ . It is easy to show that this signal can be represented exactly by a single-delay, first-order DDE:

$$\dot{x} = a_1 x + a_2 x_{\tau}. \quad (5)$$

If  $\tau$  is chosen such that  $m_{xx}(\tau) = 0$ , we find that the signal is represented by the reduced coefficients  $a_1 = 0$  and  $a_2 = -\omega$ . For our numerical analysis we generated a harmonic signal sampled at 64 points per cycle. The window length used is 10 cycles, and the time delay is  $\tau = 16$ . To train and test the detector we used 400 non-overlapping observation windows. The detector is based on the features  $a_1$  and  $a_2$  calculated with the correlation method presented in Section 2. Fig. 1 shows the receiver operating characteristic (ROC) curve for both the dynamical and a full bandwidth energy detector for a false alarm probability of  $P_{fa} = 0.1$ .



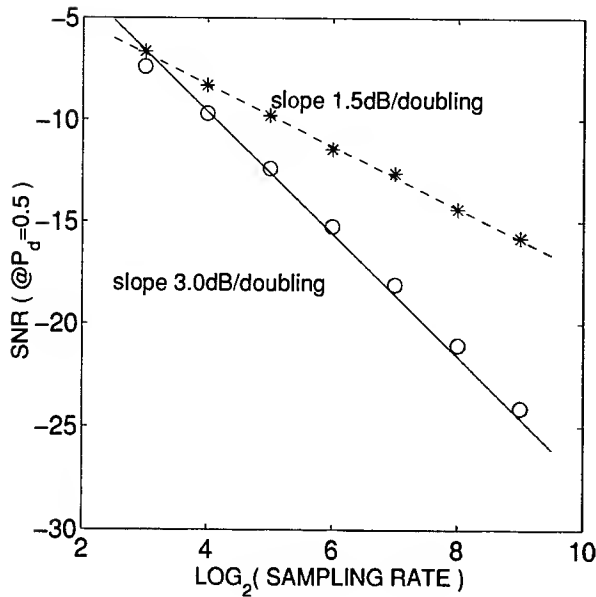


Figure 2: SNR at  $P_d = 0.5$  as a function of the sampling rate for the Rossler signal for the dynamical detector ( $\circ$ ) and energy detector (\*). The sampling rate is expressed in units of points per characteristic cycle. The slope of the dashed line has a slope of  $-1.5$  dB per factor of 2 increase in sampling rate. The solid line has a slope of  $-3.0$  dB per factor of 2 increase in sampling rate.

We note that in addition to the harmonic signal, an exponentially damped harmonic can also be represented exactly by the linear DDE model Eq. (5). In this way one can study the detection of transient pulses in the same framework.

The detection performance shown in Fig. 1 can be improved by increasing the data sampling rate, which holds true over a very broad range of signal classes [1]. This effect is tied to model specification, due to the estimate of the signal derivative. To demonstrate this we computed the ROC curves for increasing sampling rates for the  $x$ -component of the broadband, nonlinear Rossler signal [1]. The model coefficients are computed using the correlation method and a quadratic single-delay DDE [1][2].

In Fig. 2 we plot the SNR corresponding to  $P_d = 0.5$  and  $P_{fa} = 0.1$  as a function of the logarithm of the sampling rate of the Rossler signal for both the dynamical and the energy detector. We can observe a linear dependence for both detectors, however the slopes of these curves are not the same. The difference in slope accounts for a relative gain of the dynamical detector over the energy detector of about 1.5 dB per doubling of sampling rate, which is close to the theoretical gain expected from a matched filter.

These experiments show that the dynamical models can attain performance scaling close to that of a matched filter, while not requiring the original signal template. We postulate that this property is derived from the preservation of nonlinear phase relationships by the dynamical model.

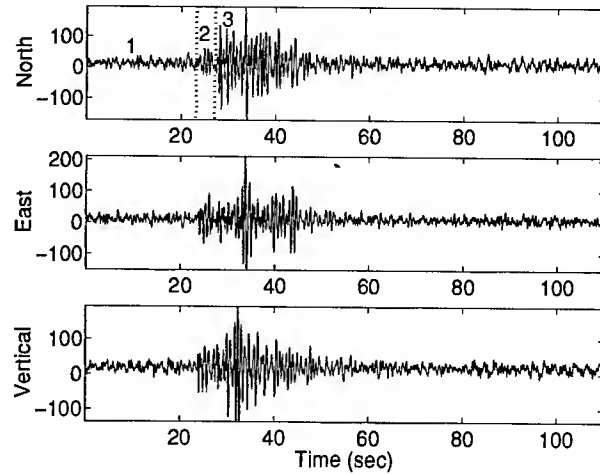


Figure 3: Typical seismic recording from an earthquake used in our data analysis. The data can be divided into three parts of interest, 1) preceding noise, 2) the P-wave, and 3) the S-wave

#### 4. CLASSIFICATION OF SEISMIC DATA

Man-made explosions around the pre-stressed area of the Panama Canal are similar in their seismic energy content to shallow earthquakes, making conventional discrimination methods difficult to use. The estimation of many seismological quantities, such as seismic hazard, seismicity, and energy release distribution is impossible with a data base polluted by man-made explosions. Therefore, it is important to discriminate between explosions and earthquakes on a routine basis. The mechanical properties of the rocks that seismic waves propagate through quickly organize the waves into two types. These are compressional waves, also known as primary or P-waves, which travel quickly, and shear waves, also known as secondary or S-waves, which travel usually at 60 to 70 percent of the speed of the P-waves. Examples of earthquake and explosion time series are shown in Fig. 3 and 4, respectively.

##### 4.1. The Data Set

The library of data consists of 20 seismic events (12 earthquakes and 8 explosions), recorded with 3-axis sensors in north, east, and vertical- direction. The data can be divided into three parts of interest (see Fig. 3), namely preceding noise, the P-wave, and the S-wave. The sampling rate is 40 Hz.

##### 4.2. Master Event Correlations

First, we give a brief description of the master event correlation analysis carried out by Persson and Boutet [8], in order to compare our results with theirs. The analysis is performed by using a library of known events to select unknown events using second, third and fourth-order cross-correlation functions. The functions are all non redundant cross-correlation combinations for each data window between the library events and the unknown events. The



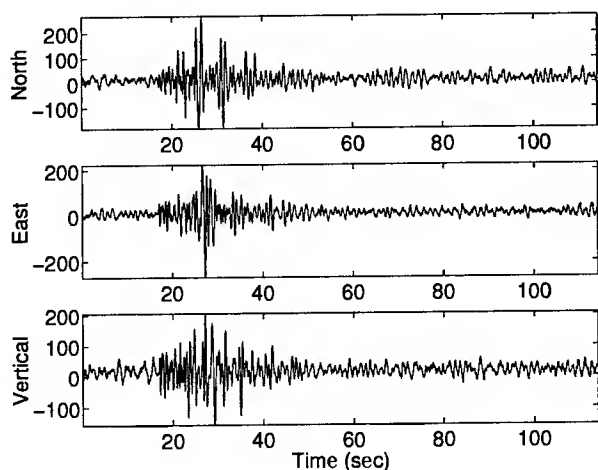


Figure 4: Typical seismic recording from an explosion used in our data analysis.

data windows are fitted to the phase of interest (i.e. noise, P-wave, or S-wave), and the lag  $\tau$  is chosen corresponding to the maximum correlation value. The second order cross-correlation between the unknown event  $x(t)$  and one of the library events  $y(t)$  is defined as,

$$m_{xy}(\tau) = \frac{1}{N} \sum_{n=\alpha}^{N+\alpha} x(t)y(t+\tau) \quad (6)$$

where  $N$  and  $\alpha$  have to be matched to the phase of interest. The third order cross-correlation is defined as,

$$m_{xxy}(\tau_1, \tau_2) = \frac{1}{N} \sum_{n=\alpha}^{N+\alpha} x(t)x(t+\tau_1)y(t+\tau_2) \quad (7)$$

The fourth order cross-correlation is defined as,

$$m_{xxxy}(\tau_1, \tau_2, \tau_3) = \frac{1}{N} \sum_{n=\alpha}^{N+\alpha} x(t)x(t+\tau_1)x(t+\tau_2)y(t+\tau_3) \quad (8)$$

The classification procedure is as follows: calculate the two master clusters composed of library events only. To classify an unknown event all cross-correlations in Eq. (6) to Eq. (8) (including correlations like  $m_{xyy}$ ,  $m_{xxyy}$ , and  $m_{xyyy}$ ) are estimated for all the library events  $y(t)$  and the unknown event  $x(t)$ , which forms the unknown cluster. The next step is to calculate the squared Mahalanobis-distance between the unknown cluster and the two master clusters. The shortest squared Mahalanobis-distance classifies the unknown event as an explosion or an earthquake. If the clusters overlap by more than 50 percent, the method is considered to have failed and the cross correlations are not used in the analysis.

To evaluate the classification performance, each of the master events are tested against the library events. The second order method discriminates 40 percent of the library events correctly, while the third and fourth order methods succeed for 75 percent and 80 percent of the library events respectively.

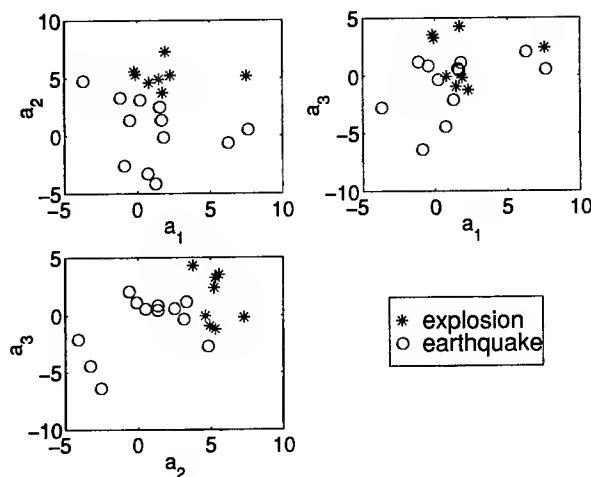


Figure 5: Parameter distribution for the library events, calculated from the P-wave in the north-direction. Note that in the two figures containing  $a_2$  the two classes are linearly separable. The processing parameters are  $i = 25$  and  $j = 19$ .

### 4.3. Nonlinear Dynamical Models

We now describe classification of the seismic data utilizing the two-delay second order DDE model given by Eq. (2). The three parameters  $a_1$ ,  $a_2$ , and  $a_3$  are estimated with Eq. (3). To solve Eq. (3) all the moments in the matrix equation have to be estimated, and we use an unbiased estimate defined as

$$\langle x^a(n)x^b(n-i)x^c(n-j) \rangle = \frac{1}{N-m} \sum_{n=m}^{N-1} x^a(n)x^b(n-i)x^c(n-j) \quad (9)$$

where  $m$  is equal to the largest of  $i$  and  $j$ ;  $N$  is the window length and  $i$  and  $j$  are the discrete delays corresponding to  $\tau_1$  and  $\tau_2$  respectively; the powers  $a$ ,  $b$  and  $c$  are set to 0, 1 or 2 corresponding to the moment that has to be calculate. We use the window length  $N = 128$  samples, and the two delays  $i$  and  $j$  are set to 25 and 19 respectively, which give the largest possible  $a_3$  coefficient. The feature space spanned by the three model coefficients is shown in Fig. 5. One can observe that the explosions and earthquakes are linearly separable. Moreover, the signal separation is significant from noise signals as well (not shown here).

To achieve discrimination between the earthquake and explosion recordings, we develop classifiers based on the above theory which successfully separates earthquake and explosion time series. For a quantitative analysis of the classification performance we apply both Mahalanobis-distance decision criteria and a neural network algorithm to discriminate between the library events based on their model representation in the feature space. In order to compare the dynamical model approach to the method of master-event correlation we implement a Mahalanobis-distance classifier of the same type as the classifier used by Persson and Boutet [8]. The classification performance achieved with the classi-

Direction	North		Vertical		East	
Wave	s	p	s	p	s	p
Correct Earthquake Classification	75	92	67	67	92	83
Correct Explosion Classification	75	100	88	88	75	75

Table 1: Classification of the library events with the dynamical model combined with Mahalanobis-distance decision criteria. All values in the table are given in percent.

Direction	North		Vertical		East	
Wave	s	p	s	p	s	p
Correct Earthquake Classification	69	88	91	88	100	96
Correct Explosion Classification	78	98	75	70	86	74

Table 2: Classification of the library events with the dynamical model combined with a LVQ neural net. All values in the table are given in percent.

fier based on the dynamical model combined with Mahalanobis-distance decision criteria for various wave types (S or P) and directions (North, East, Vertical) are presented in Table 1. The best overall result, 96 percent correct classification, is achieved for the P-wave in the north-direction. Next we build a neural net classifier where the model parameters are used as input to a learning vector quantization (LVQ) neural net. LVQ is a method for training competitive layers in a supervised manner. To evaluate the classification ability we train the LVQ neural net to discriminate between earthquakes and explosions on all but one of the library events, and then use the trained neural net to declare the removed event as an explosion or an earthquake. This is performed for all library events, in all possible combinations in terms of phase and direction. The classification ability of a LVQ neural net can be slightly different between trainings, even if the training is repeated on the exact same data set. In order to reduce these type of fluctuations the training and classification procedure is repeated 10 times and then the average classification performance is calculated. The results for various wave types (S or P) and directions (North, East, Vertical) are summarized in Table 2. The best overall result, 93 percent correct classification, is achieved for the P-wave in the north-direction.

## 5. CONCLUSIONS

We have developed a method for the estimation of DDE models motivated by the Yule-Walker equations, which provides computational speed, numerical stability and noise robustness. With numerical experiments we showed that a

detector based on these dynamical models can achieve scaling close to that of a matched filter, without requiring the original signal template.

Nonlinear dynamical signal models can be utilized for detection and classification of a wide range of signals. In this work we analyzed real-world seismic signals derived from seismic station recordings in the region of the Panama Canal, which are of a transient nature. We implemented two classifiers based on the two-delay second order DDE model given by Eq. (2). The classifier that uses the Mahalanobis-distance decision criteria results in 96 percent correct classification of the library events. The classifier built with a LVQ neural net results in 93 percent correct classification.

We compare our results with a previous classification method used on the same seismic recording database, which used time-domain master-event correlations of second, third, and fourth order. The best classification performance achieved with the master-event correlations is 80 percent for the fourth order. Hence, we find improved performance over even the highest-order master-event correlations method, which indicates that a dynamical model approach incorporates additional signal information on this example. In summary, this example is important because it indicates that dynamical modeling and classification methods can add additional performance gain in a real-world setting.

## REFERENCES

- [1] Kadtke J., Pentek A., *Automated signal classification using dynamical signal models and generalized higher-order data correlations (U)*, USN Journal of Underwater Acoustics, in press (2000).
- [2] Kadtke J., Kremliovsky M., *Estimating dynamical models using generalized moment functions*, Physics Letters A **260**, 203 (1999).
- [3] Hale, J.K., Lunel, S.M.V., *Introduction to functional differential equations*, Springer-Verlag, 1993.
- [4] Voss, H., Kurths, J., *Reconstruction of non-linear time delay models from data by the use of optimal transformations*, Physics Letters A **234**, 336-44, 1997.
- [5] Boashash, B., Edward J.P., Abdelhak, M.Z., eds., *Higher-order statistical signal processing*, Longman and Wiley Press, Melbourne, 1995.
- [6] Kremliovsky, M., Kadtke, J., Inchiosa, M., Moore, P., *Characterization of dolphin acoustic echo-location data using a dynamical classification method*, International Journal of Bifurcation and Chaos **8**, 813-23, 1998.
- [7] Lennartsson R. K., *Classification with dynamical models estimated with higher order statistical moments*, FOA Report, FOA-R-99-01292-313-SE, National Defence Research Establishment, Sweden, November 1999.
- [8] Persson L., Boutet, J., *Discrimination of local seismic events in panama by means of higher-order statistics*, Proceedings of IEEE Signal Processing Workshop on Higher-Order Statistics, Banff, Alberta, Canada, July 21-23, 1997, pp. 14-19.
- [9] Marple, S.L., *Digital spectral analysis*, Englewood Cliffs, NJ, Prentice Hall, 1987.

# THE PERFORMANCE OF SPARSE TIME-REVERSAL MIRRORS IN THE CONTEXT OF UNDERWATER COMMUNICATIONS

*João Gomes*

*Victor Barroso*

Instituto Superior Técnico – Instituto de Sistemas e Robótica

Av. Rovisco Pais, 1049-001 Lisboa, Portugal

{jpg,vab}@isr.ist.utl.pt

## ABSTRACT

Recently, wave focusing using a uniform time-reversal array has been demonstrated in the ocean with very encouraging results. This technique may be used to regenerate a mildly distorted signal at the input of a digital underwater acoustic receiver, hence reducing its equalization requirements at the expense of additional complexity at the transmitter. This work investigates the performance improvements that become possible when sparse and non-uniform arrays are used. Results from the theory of randomly-spaced arrays are extended to a simplified ocean waveguide, revealing that familiar relations between the sensor placement density function and the directional characteristics of the generated acoustic field are still valid in the large-scale. Simulation results confirm the validity of these derivations.

## 1. INTRODUCTION

Underwater acoustic propagation is a waveguide phenomenon where pressure waves are repeatedly reflected by the sea surface and bottom, and undergo time-varying refraction and scattering by inhomogeneities in the medium [1]. When acoustic waves are used to transmit digitally modulated signals, these physical processes induce temporal spreading of the signaling waveforms, resulting in significant intersymbol interference (ISI) upon reception. Reliable decoding of (relatively) high-speed phase-coherent modulations under such conditions requires the use of spatial diversity and powerful, computationally intensive, equalization algorithms [2].

A wave-focusing approach is used in this work to mitigate the effects of ISI, so that the receiver may be simplified. Focusing waves in inhomogeneous media is a difficult problem that usually requires detailed physical knowledge of the environment. Severe performance

degradation may occur if the assumed propagation conditions do not match the actual ones to a high degree of accuracy that is unattainable under most realistic conditions in the ocean. As an alternative to open-loop operation, wave focusing through phase conjugation may be used whenever the desired focal point has the ability to generate energy, either by active means, or by reflection of incoming waves.

Phase conjugation in underwater acoustics is implemented through a time-reversal mirror (TRM), i.e., an array of transducers that record sound, store it, and later reproduce it backwards in time [3, 4]. The generated waves propagate in a manner reciprocal to the original field, such that energy is automatically redirected towards the focus and concentrated there even when poorly characterized regions are crossed. An application of phase conjugation to underwater communication requires the receiver to first transmit the basic waveforms of the signal constellation, so that their distorted replicas are stored at the mirror. These are subsequently used to modulate a message, regenerating a nearly multipath-free signal with the desired pulse shape at the focus, where the receiver is located [5, 6].

In [7] it was shown experimentally that a time-reversal mirror may still perform adequately even when the sensors are spaced by about ten wavelengths. Moreover, it is known that nonuniform sensor separation may dramatically improve the capacity of large linear arrays for direction of arrival estimation [8, 9]. Motivated by these results, the goal of the present paper is to study sparsely-populated mirrors and sensor allocation strategies that make efficient use of the available degrees of freedom.

## 2. DATA MODEL

For the sake of analytical tractability, the ocean waveguide is modeled as a range-independent cross-section with depth  $H$  and constant sound speed  $c$ . The ocean surface is an ideal pressure-release surface, while the

This work was partially funded by the EU under MAST project ASIMOV, and by FEDER and PRAXIS XXI, under project INFANTE.

bottom is rigid and lossy. Their (constant) reflection coefficients are  $-1$  and  $0 < \alpha_B < 1$ , respectively.

From a linear systems perspective, the normalized transfer function (Green's function) at frequency  $\omega$  from range/depth  $\mathbf{r}' = (R', z')$  to point  $\mathbf{r}$  is obtained by solving the wave equation for a time-harmonic point source

$$[\nabla^2 + k^2(\mathbf{r})] G_\omega(\mathbf{r}, \mathbf{r}') = -\delta(\mathbf{r} - \mathbf{r}'), \quad (1)$$

where  $k(\mathbf{r}) = \omega/c(\mathbf{r})$  is the wavenumber. If the medium is bounded,  $G_\omega$  must satisfy appropriate boundary conditions. In this work it is assumed that frequencies of several KHz are used, in which case ray theory provides adequate modeling of the acoustic propagation. The solution of (1) may then be approximated by a series of eigenray contributions

$$G_\omega(\mathbf{r}, \mathbf{r}') \approx \sum_i (-1)^{n_{S,i}} \alpha_B^{n_{B,i}} \{a_i \exp(-j\omega\tau_i)\}, \quad (2)$$

where  $n_{S,i}$ ,  $n_{B,i}$  are the number of surface and bottom reflections of the  $i$ -th eigenray linking  $\mathbf{r}'$  and  $\mathbf{r}$ . Acoustic rays travel in straight lines under the assumed isovelocity conditions, in which case the term inside  $\{\cdot\}$  in (2) may be expressed as a free-space Green's function linking  $\mathbf{r}'$  and a point  $\mathbf{r}^{(i)}$  whose distance from  $\mathbf{r}'$  equals the length of the  $i$ -th eigenray

$$G'_\omega(\mathbf{r}^{(i)}, \mathbf{r}') = -\frac{\exp(-jk|\mathbf{r}^{(i)} - \mathbf{r}'|)}{4\pi|\mathbf{r}^{(i)} - \mathbf{r}'|} \quad (3)$$

Reciprocity of the medium allows the spatial arguments of  $G_\omega(\cdot, \cdot)$  and  $G'_\omega(\cdot, \cdot)$  to be interchanged.

Propagation from a source to each transducer of the TRM and back generates an intricate pattern of eigenrays. Under homogeneous conditions, decomposition of (2) as a weighted sum of free-space terms leads to the image method [1], where points in the water column are expanded into a series of surface- and bottom-reflected images. The original waveguide problem is thus transformed into one of (coupled) free-space propagation between the source and the image arrays (fig. 1), which provides more insight into the operation of the TRM.

Let  $\mathbf{r}_a = (R_a, z_a)$  be a convenient reference point for the array. Henceforth, unless otherwise noted, displacements will be relative to  $\mathbf{r}_a$ . Additional coordinate systems will be placed at all  $(p, m)$  images of  $\mathbf{r}_a$  (see fig. 1,) and their  $z$  axis oriented so that the coordinates of the associated virtual sensors are independent of  $(p, m)$ . When expressed in frame  $(p, m)$ , the displacement to an arbitrary field point  $\mathbf{r}$  is denoted by  $\mathbf{r}^{(p,m)}$ . It is then possible to write (2) in vector form, discarding contributions from rays that suffer more than  $N_B$  bottom reflections

$$G_\omega(\mathbf{r}, \mathbf{r}') = \begin{bmatrix} \alpha \\ -\alpha \end{bmatrix}^H \begin{bmatrix} G'_\omega^{(0)}(\mathbf{r}, \mathbf{r}') \\ G'_\omega^{(1)}(\mathbf{r}, \mathbf{r}') \end{bmatrix} \quad (4)$$

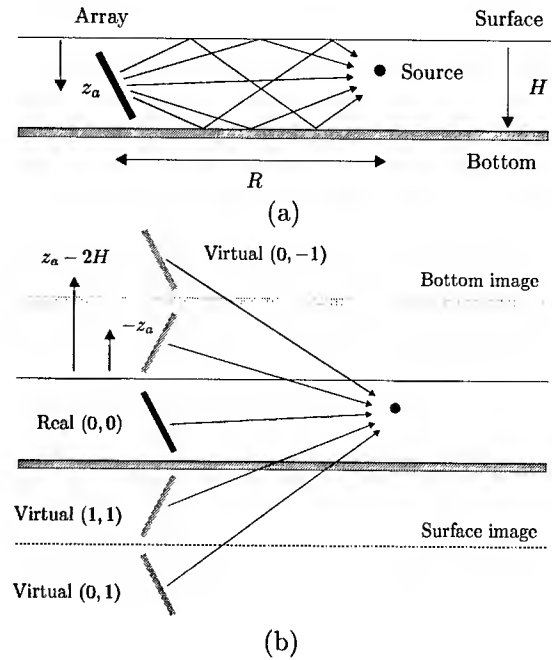


Figure 1: Field computation by the image method (a) Waveguide propagation (b) Image expansion

$$\alpha = [(-\alpha_B)^{|-N_B|} \dots (-\alpha_B)^{|N_B|}]^T$$

$$G'_\omega^{(p)}(\mathbf{r}, \mathbf{r}') = \begin{bmatrix} G'_\omega(\mathbf{r}, \mathbf{r}'^{(p, -N_B)}) \\ \vdots \\ G'_\omega(\mathbf{r}, \mathbf{r}'^{(p, N_B)}) \end{bmatrix}, p = 0, 1.$$

### 3. TIME-REVERSAL MIRROR

Time reversal of a real signal is equivalent to conjugation of its Fourier transform. Then, the normalized field produced by a TRM due to a point source at  $\mathbf{r}_s$  is obtained by summing the sensor contributions as follows

$$P_\omega(\mathbf{r}, \mathbf{r}_s) = \sum_m G_\omega^*(\mathbf{r}_m, \mathbf{r}_s) G_\omega(\mathbf{r}, \mathbf{r}_m)$$

$$= \alpha^H \left[ \mathbf{B}_\omega^{(0,0)}(\mathbf{r}, \mathbf{r}_s) + \mathbf{B}_\omega^{(1,1)}(\mathbf{r}, \mathbf{r}_s) - 2\text{Re} \left\{ \mathbf{B}_\omega^{(0,1)}(\mathbf{r}, \mathbf{r}_s) \right\} \right] \alpha, \quad (5)$$

where, by reciprocity, the beamforming matrices  $\mathbf{B}_\omega$  are given by

$$\mathbf{B}_\omega^{(p,q)}(\mathbf{r}, \mathbf{r}_s) = \sum_m G'_\omega^{(p)}(\mathbf{r}_m, \mathbf{r}) G'^{(q)H}(\mathbf{r}_m, \mathbf{r}_s). \quad (6)$$

Now the source is assumed to be in the far-field of each array image, so that a plane wave approximation for the free-space Green's function (3) may be used

$$G'_\omega(\mathbf{r}, \mathbf{r}') \approx -\frac{\exp(-jkr)}{4\pi r} \exp(jk\langle \mathbf{r}', \mathbf{e}_r \rangle), \quad |\mathbf{r}'| \ll |\mathbf{r}|,$$

where  $r = |\mathbf{r}|$ ,  $\mathbf{e}_r = \mathbf{r}/r$ , and  $\langle \cdot, \cdot \rangle$  denotes the inner product of two vectors. Each element of the beamforming matrices (6) has the form

$$B_{\omega m,n}^{(p,q)} = C_{\omega}(r^{(p,m)}, r_s^{(q,n)}) D_{\omega}(\mathbf{e}_r^{(p,m)} - \mathbf{e}_s^{(q,n)}) \quad (7)$$

$$C_{\omega}(r, r') = \frac{\exp(-jk(r - r'))}{(4\pi)^2 r r'} \quad (8)$$

$$D_{\omega}(\mathbf{e}) = \sum_m \exp(jk \langle \mathbf{r}_m, \mathbf{e} \rangle). \quad (9)$$

In (9),  $D_{\omega}$  is recognized as an array directivity function. When  $\mathbf{e}_r^{(p,m)} = \mathbf{e}_s^{(q,n)}$  each term in the sum is equal to unity, and the contributions from all elements add in phase in this direction. In other directions, the contributions are not in phase, and the field is smaller.

In (5), both  $\mathbf{B}_{\omega}^{(0,1)}$  and the off-diagonal terms of  $\mathbf{B}_{\omega}^{(0,0)}$  and  $\mathbf{B}_{\omega}^{(1,1)}$  account for the influence at  $\mathbf{r}^{(p,m)}$  of a beampattern steered toward  $\mathbf{r}_s^{(q,n)}$ . Field calculations then require evaluating the influence of each virtual array on all images of the target ocean section. The total field  $P_{\omega}$  is obtained by a (weighted) sum over array images and target images. If the aperture and sensor density are large enough so that  $D_{\omega}$  is narrow and has a single main lobe, image arrays only contribute significantly in the main (0,0) cross-section. In that case, the mirror operates in retrodirective mode by sending acoustic beams in the same directions where it receives energy from the source during the first transmission. The large-scale shape of the focal region is mostly determined by the beampattern  $D_{\omega}$ , while the fine-scale structure results from the interference of beams, and is heavily influenced by  $C_{\omega}$ .

#### 4. RANDOMLY-SPACED SENSORS

The theory of randomly-spaced arrays in free space shows that the beamwidth depends mainly on the aperture dimension, while the directive gain and sidelobe level are directly related to the number of elements used if the average spacing is large [8]. These results will now be extended to the case where multiple array images exist.

The array is formed by  $M$  vertically-placed sensors at  $\mathbf{r}_m = (0, z_m)$  relative to the reference point  $\mathbf{r}_a$ . The depths  $\{z_m\}$  are assumed to be i.i.d. random variables with a common probability density function that is greater than zero only in an interval of length  $L$ . According to (9), the radiation pattern is

$$D(u) = \sum_{m=1}^M \exp(jux_m), \quad (10)$$

where  $x_m = \frac{2}{L} z_m$  is the normalized sensor depth and  $u = k \frac{L}{2} \langle (0, 1), \mathbf{e} \rangle$  may be interpreted as a scaled direction cosine. The pdf of  $x$  will be denoted by  $g(x)$ , with characteristic function  $\phi(u) = \exp(jux)$ . In the assumed array vertical geometry the argument  $u$  already incorporates the acoustic frequency through the wavenumber  $k$ , hence the explicit dependence of  $D$  on  $\omega$  is dropped in (10).

It follows from the definition of the characteristic function that the mean beampattern in (10) is

$$E\{D(u)\} = \sum_{m=1}^M E\{\exp(jux)\} = M\phi(u). \quad (11)$$

Since the phase-conjugated field (5) only depends on the sensor positions through  $D$ , it is clear that the mean acoustic pressure  $\bar{P}_{\omega}(\mathbf{r}, \mathbf{r}_s)$  is given by a similar expression, with  $D$  replaced by  $M\phi$  in the beamforming matrices, as shown by

$$\bar{B}_{\omega m,n}^{(p,q)} = C_{\omega}(r^{(p,m)}, r_s^{(q,n)}) \cdot M\phi(u) \quad (12)$$

$$u = k \frac{L}{2} (\sin \theta_r^{(p,m)} - \sin \theta_s^{(q,n)}), \quad (13)$$

where  $\theta_r^{(p,m)}$  and  $\theta_s^{(q,n)}$  are the bearing angles to the field and source images, respectively. As in the free-space case, the mean field is identical to the one that would be created by a continuous aperture with excitation  $g(x)$ .

The variance of the time-reversed field is denoted by  $\sigma^2(\mathbf{r}, \mathbf{r}_s) = E\{|P_{\omega}(\mathbf{r}, \mathbf{r}_s) - \bar{P}_{\omega}(\mathbf{r}, \mathbf{r}_s)|^2\}$ , and involves evaluating the sum

$$\begin{aligned} \sigma^2(\mathbf{r}, \mathbf{r}_s) &= \sum_{i,l,m,n,p,q,u,v} (-1)^{(p-q)-(u-v)} \alpha_i \alpha_l \alpha_m \alpha_n \\ &\quad \times E\{\Delta B_{\omega m,n}^{(p,q)} \Delta B_{\omega i,l}^{(u,v)*}\} \\ \Delta B_{\omega m,n}^{(p,q)} &= B_{\omega m,n}^{(p,q)} - \bar{B}_{\omega m,n}^{(p,q)} \\ &= C_{\omega}(r^{(p,m)}, r_s^{(q,n)}) (D(u) - M\phi(u)), \end{aligned} \quad (14)$$

where the argument  $u$  is defined in (13). Given the i.i.d. assumption on sensor positions, the free-space covariance function satisfies [8]

$$\begin{aligned} E\{(D(u) - M\phi(u))(D(v) - M\phi(v))^*\} &= \\ M(\phi(u - v) - \phi(u)\phi^*(v)) &\approx M\phi(u - v), \end{aligned} \quad (15)$$

for sufficiently large  $u, v$ . Using (8), (14) and (15), the terms of  $\sigma^2(\mathbf{r}, \mathbf{r}_s)$  are seen to depend mostly on argument differences

$$\begin{aligned} E\{\Delta B_{\omega m,n}^{(p,q)} \Delta B_{\omega i,l}^{(u,v)*}\} &= \\ \frac{\exp(-jk((r^{(p,m)} - r_s^{(q,n)}) - (r^{(u,i)} - r_s^{(v,l)})))}{(4\pi)^4 r^{(p,m)} r_s^{(q,n)} r^{(u,i)} r_s^{(v,l)}} \end{aligned}$$

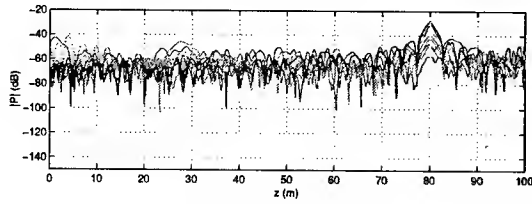


Figure 2: ULVA performance

$$\times M \phi(k \frac{L}{2} ( \sin \theta_r^{(p,m)} - \sin \theta_s^{(q,n)} ) - ( \sin \theta_r^{(u,i)} - \sin \theta_s^{(v,l)} ) ) . \quad (16)$$

More importantly, both (16) and  $\sigma^2(\mathbf{r}, \mathbf{r}_s)$  depend on the number of array sensors only through the linear term  $M$ . This shows that, for a given physical configuration and placement pdf, the variance of the normalized beampattern  $\frac{P_w}{M}$  decreases to zero with  $\frac{1}{M}$ , as in free-space. For large  $M$ , individual mirror responses will therefore be close to the average pressure  $\bar{P}_w$  with high probability.

## 5. SIMULATION RESULTS

The parameters that were used in the simulations are listed in table 1. Both uniform and square-cosine densities were considered as sensor placement strategies [8], with nonzero support in the depth interval [1, 99] m. The expressions for their densities and characteristic functions are shown in table 2. Figure 2 shows the expected mirror performance evaluated using (5) for uniform linear vertical arrays (ULVA) with  $M = 5, 10, 15, 20, 30, 50, 70$ , and 100 sensors, evenly-spaced between 1 m and 99 m. The focusing effect is still clearly visible when 50 sensors are used ( $5.3\lambda$  spacing,) but becomes severely degraded for lower values of  $M$ . Figure 3 shows the average time-reversed acoustic field for the densities of table 2, evaluated using (5) with asymptotic beamforming matrices (12). The corresponding free-space responses  $\phi(u)$  are also superimposed on these plots. Several curves are shown for different values of  $M$  to simplify the comparison with Monte Carlo simulations,

Table 1: Simulation parameters

Bottom depth	$H = 100$ m
Source range	$R = 1500$ m
Source depth	$z_s = 80$ m
Frequency	$f = 4$ KHz
Sound speed	$c = 1500$ ms <sup>-1</sup>
Bottom reflectivity	$\alpha_B = 0.3$
Image truncation limit	$N_B = 10$

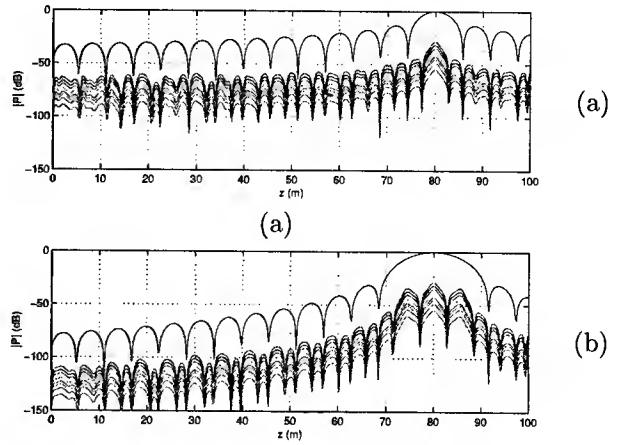


Figure 3: Expected field with random spacing (a) Uniform pdf (b) Square-cosine pdf

although it is clear from the previous discussion that  $M$  only introduces a gain in the mean field. These results confirm that the large-scale evolution of the field is determined by  $\phi(u)$ , although the detailed behavior depends on the interference pattern between array images. In particular, the acoustic field between the pressure nulls at 77 m and 83 m is almost identical in figures 2 and 3.

The time-reversed field of figure 3b seems to be more suitable for coherent communication applications, since it creates a broader region of high acoustic energy around the focus. As shown in [5], the extent of the low ISI zone may be estimated by considering the joint evolution of the acoustic field for the higher and lower frequencies in the PAM signaling pulses. From that perspective, concentrating energy in a broad main lobe maximizes the region where the monochromatic field components within the signal bandwidth behave coherently, leading to low spectral pulse distortion and mild ISI.

Figure 4 shows the average acoustic fields that were obtained for the previously considered values of  $M$  in 500 Monte Carlo simulations. The results are in good agreement with the theoretical mean values of figure 3, even for the lowest values of  $M$ . The difference in residual sidelobe level may be attributed to model discrepancies, since the ideal responses of figure 3 were based

Table 2: Densities and characteristic functions

	Uniform	Square-Cosine
$g(x),  x  < 1$	$\frac{1}{2}$	$\cos^2 \frac{\pi x}{2}$
$\phi(u)$	$\frac{\sin u}{u}$	$\frac{\sin u}{u} \frac{1 - (\frac{u}{\pi})^2}{1 - (\frac{u}{\pi})^2}$

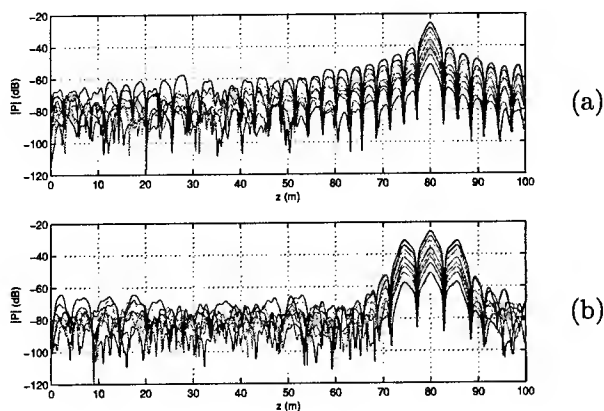


Figure 4: Average field in Monte Carlo simulations (a) Uniform pdf (b) Square-cosine pdf

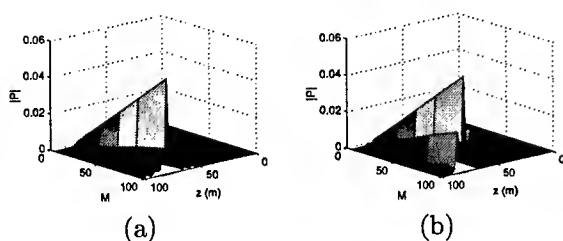


Figure 5: Average field evolution with the number of sensors (a) Uniform pdf (b) Square-cosine pdf

on a plane wave approximation. Naturally, individual time-reversed fields vary considerably, especially when few sensors are used, but beampatterns with globally desirable features are obtained with reasonably high probability for  $M > 30$ . Closer examination of the field covariance is deferred to future work.

The same results of figure 4 are represented in figure 5 using a linear scale, showing that the mean field does indeed increase linearly with  $M$ .

## 6. CONCLUSION

Non-uniform sensor placement strategies for linear time-reversal arrays were investigated as a means of (i) reducing the required number of elements relative to uniform geometries, and (ii) assessing the sensitivity of focusing power to sensor locations.

Using a ray propagation model, results from the theory of antenna arrays with randomly-spaced elements in free space were extended to the ocean waveguide, that is crucial for effective operation of a time-reversal mirror. Simulations have confirmed the validity of theoretical predictions for two distinct sensor placement distributions. The results indicate that focusing performance is affected mainly by the total array

length, rather than inter-sensor separation or precise sensor locations. Good results are obtained even when the average spacing of sensors is significantly larger than half a wavelength, since grating lobes are not coherently combined. These conclusions are supported by the work of other authors using uniform arrays [10].

While these preliminary studies indicate that some reduction in the number of sensors is possible relative to uniform arrays without incurring significant performance losses, practical considerations prevent randomly-spaced mirrors from attaining the spectacular savings envisaged in [8] for arrays with several thousand elements.

## 7. REFERENCES

- [1] F. Jensen *et al.*, *Computational Ocean Acoustics*, AIP Press, 1994.
- [2] M. Stojanovic, "Recent advances in underwater acoustic communications," *IEEE J. of Oceanic Eng.*, vol. 21, no. 2, pp. 125–136, Apr. 1996.
- [3] D. Jackson, D. Dowling, "Phase conjugation in underwater acoustics," *J. Acoust. Soc. Am.*, vol. 89, no. 1, pp. 171–181, Jan. 1991.
- [4] W. Kuperman *et al.*, "Phase conjugation in the ocean: Experimental demonstration of an acoustic time-reversal mirror," *J. Acoust. Soc. Am.*, vol. 103, no. 1, pp. 25–40, Jan. 1998.
- [5] J. Gomes, V. Barroso, "Using phase conjugation for underwater acoustic communication: Design guidelines," in *Proc. EUSIPCO 2000*, Tampere, Finland, Sept. 2000.
- [6] J. Gomes, V. Barroso, "Ray-based analysis of a time-reversal mirror for underwater acoustic communication," in *Proc. ICASSP 2000*, Istanbul, Turkey, June 2000.
- [7] W. Kuperman *et al.*, "Ocean acoustic time-reversal mirror," in *Proc. 4th European Conf. on Underwater Acoustics*, Rome, Italy, 1998, pp. 493–498.
- [8] Y. Lo, "A mathematical theory of antenna arrays with randomly spaced elements," *IEEE Trans. on Antennas and Propag.*, vol. AP-12, no. 3, pp. 257–268, May 1964.
- [9] C. Chambers *et al.*, "Temporal and spatial sampling influence on the estimates of superimposed narrow-band signals: When less can mean more," *IEEE Trans. on Sig. Proc.*, vol. 44, no. 12, pp. 3085–3098, Dec. 1996.
- [10] A. Abrantes, "Examination of time-reversal acoustics in shallow water and applications to underwater communications," M.Sc. thesis, Naval Postgraduate School, Monterey, CA, June 1999.



# BEAM PATTERNS OF AN UNDERWATER ACOUSTIC VECTOR HYDROPHONE†

Kainam Thomas WONG  
*Department of Electronic Engineering,  
 Chinese University of Hong Kong,  
 Shatin, NT, Hong Kong  
 ktwong@ieee.org*

Hoiming CHI  
*School of Electrical & Computer Engineering,  
 Purdue University,  
 West Lafayette, IN 47907-1285 U.S.A.  
 hoiming@purdue.edu*

## ABSTRACT

A vector hydrophone is composed of two or three spatially collocated but orthogonally oriented velocity hydrophones plus an optional collocated pressure hydrophone. A vector hydrophone may form azimuth-elevation beams that are frequency invariant, bandwidth invariant and same for the near field as for the far field. This paper characterizes the maximum-SINR beam pattern and the matched-filter beam pattern associated with a single underwater acoustic vector hydrophone.

## 1. VECTOR HYDROPHONE

A vector hydrophone consists of two or three orthogonally oriented velocity hydrophones plus an optional pressure hydrophone, all spatially collocated in a point-like geometry. Each velocity hydrophone has intrinsic directional response to the incident underwater acoustic particle velocity wavefield, measuring one Cartesian component of the three-dimensional particle velocity vector of the incident wavefield. On the other hand, a pressure hydrophone measures the acoustical pressure as a scalar entity. A single vector hydrophone thus has an intrinsic two-dimensional azimuth-elevation directivity that is independent of signal frequency, signal bandwidth, and the source's location in the near field as opposed to the far field. In contrast, the directivity obtainable from an array of spatially displaced pressure hydrophones is based on the frequency dependent inter-hydrophone spatial phase factor; and the beam pattern consequently depends on the signal frequency and the signal bandwidth.

Velocity hydrophone technology has been used in underwater acoustics for some time [1] and currently attracts re-invigorated attention [18]. Many different types of vector hydrophones have been implemented (see the references cited in [13]). The Swallow floats [7], a freely drifting array of vector hydrophones, are neutrally buoyant and may be ballasted to any desired depth in the ocean. The DIFAR array [9] is a uniform vertical array with acoustic band limit of 270Hz for linearly constrained minimum variance beamforming with given angles and with flux gate compasses to measure the orientation of the horizontal velocity hydrophones. D'Spain, Hodgkiss & Edmonds [8] develop a vector hydrophone for infrasonic frequencies from 1 to 20 Hz.

Nehorai & Paldi [13] first develop the measurement model of the vector hydrophone and introduced it to the signal processing research community. The use of the vector hydrophone in sensor array direction finding has been inves-

tigated in [13-21, 25, 27]. Vector-hydrophone Capon spectrum estimation along pre-determined spatial direction has been investigated by D'Spain, Hodgkiss & etc. [9] and Hawkes & Nehorai [23]; and a few very compact expressions of vector hydrophone matched-filter beam pattern have been derived in [23]. However, no detailed analysis is therein provided; and vector hydrophone beam pattern analysis remains largely overlooked in the open literature.

The present work characterizes and contrasts in detail the maximum signal-to-interference-and-noise (SINR) beam patterns and the matched-filter beam patterns for each of the following possible vector hydrophone constructions.

**Construction #1:** Three orthogonally oriented velocity hydrophones plus a pressure hydrophone, in spatial collocation, give a  $4 \times 1$  array manifold [13]:

$$\mathbf{a}_k = \mathbf{a}_k^{(3+1)} \stackrel{\text{def}}{=} \begin{bmatrix} \sin \theta_k \cos \phi_k \\ \sin \theta_k \sin \phi_k \\ \cos \theta_k \\ 1 \end{bmatrix} = \begin{bmatrix} u(\theta_k, \phi_k) \\ v(\theta_k, \phi_k) \\ w(\theta_k) \\ 1 \end{bmatrix} \quad (1)$$

The first, second and third component above correspond respectively to the velocity-hydrophone aligned along respectively the x-axis, the y-axis and the z-axis; these first three components of  $\mathbf{a}_k$  give the three Cartesian direction-cosines. The last component corresponds to the pressure-hydrophone. The Frobenius norm of the first three components of any source's array manifold always equals to unity, regardless of source parameters. With this four-component vector-hydrophone construction, sources may be located to either side of the array; that is,  $\theta_k$  may range from  $0 \leq \theta_k < \pi$  instead of  $0 \leq \theta_k < \pi/2$ . The presence of the pressure hydrophone helps to distinguish between acoustic compressions and dilations. This is important because acoustic particle motion sensors (such as a velocity hydrophone), by themselves, suffer a  $180^\circ$  ambiguity, with their plane-wave response given by the "figure 8" curve. However, the addition of a pressure hydrophone breaks this ambiguity because a hydrophone distinguishes between acoustical compressions and dilations.

**Construction #2:** Three orthogonally oriented velocity hydrophones give a  $3 \times 1$  array manifold:

$$\mathbf{a}_k = \mathbf{a}_k^{(3+0)} \stackrel{\text{def}}{=} \begin{bmatrix} \sin \theta_k \cos \phi_k \\ \sin \theta_k \sin \phi_k \\ \cos \theta_k \end{bmatrix} \quad (2)$$

**Construction #3:** Two orthogonally and horizontally oriented velocity hydrophones plus a pressure hydrophone

†This research work was supported by the Hong Kong Research Grant Council's Mainline Research Grant no. 44M5010 and Direct Grant no. 2050187.



give a  $3 \times 1$  array manifold:

$$\mathbf{a}_k = \mathbf{a}_k^{(2+1)} \stackrel{\text{def}}{=} \begin{bmatrix} \sin \theta_k \cos \phi_k \\ \sin \theta_k \sin \phi_k \\ 1 \end{bmatrix} \quad (3)$$

This suffices to completely characterize the underwater acoustical velocity-field, despite the absence of the z-axis velocity hydrophone. The omission of the vertical velocity-hydrophone avoids direct measurement of the vertical component of the underwater acoustical particle motion, thereby allowing actual ocean acoustics to be better modeled as rectilinear. Because, particle motion may be circularly and elliptically polarized and needs not be rectilinear. Even if the source initially generates a single plane wave, the multipath propagation properties of the ocean environment typically lead to elliptically polarized particle motion. That there exists no vertically oriented velocity hydrophone means that non-rectilinear motion will affect the measured data minimally; and the rectilinear data model will better fit generally non-rectilinear ocean acoustics. An example of construction #3 is the cardioid [11].

**Construction #4:** Two orthogonally and horizontally oriented velocity hydrophones give a  $2 \times 1$  array manifold:

$$\mathbf{a}_k = \mathbf{a}_k^{(2+0)} \stackrel{\text{def}}{=} \begin{bmatrix} \sin \theta_k \cos \phi_k \\ \sin \theta_k \sin \phi_k \end{bmatrix} \quad (4)$$

## 2. VECTOR HYDROPHONE BEAMFORMING

A transmitting sensor array beamformer focuses its transmission energy towards targeted azimuth-elevation angular sectors, whereas a receiving sensor array beamformer represents a spatial filtering operation to separate the desired signals from interferences and noise based on their different arrival angles. A receiving beamformer may be classified as data independent or statistically optimal [6]. The former effects an a priori specified spatial angular beamformer response independent of the incoming data, but the latter adaptively optimizes certain statistical criterion defined with respect to the collected data. A matched-filter beamformer is an example of the former; and a maximum-signal-to-interference-plus-noise-ratio (maximum-SINR) beamformer is an example of the latter.

To facilitate subsequent discussion, define  $\psi$  as the angle between the two steering vectors  $\mathbf{a}_s^{(3+0)}$  and  $\mathbf{a}_i^{(3+0)}$  (each of which contains as components the Cartesian direction cosines,

$$\begin{aligned} \cos \psi &\stackrel{\text{def}}{=} \frac{(\mathbf{a}_i^{(3+0)})^H \mathbf{a}_s^{(3+0)}}{\|\mathbf{a}_s^{(3+0)}\| \|\mathbf{a}_i^{(3+0)}\|} \\ &= \sin \theta_s \sin \theta_i \cos(\phi_s - \phi_i) + \cos \theta_s \cos \theta_i \end{aligned}$$

### 2.1. Matched-Filter Beamforming

Matched-filter beamforming forms a data-independent beamforming weight vector  $\mathbf{w}_{MF}$  to match the desired signal's steering vector  $\mathbf{a}(\theta_s, \phi_s)$ ; that is,  $\mathbf{w}_{MF} = \mathbf{a}(\theta_s, \phi_s)$ . With the array nominally pointing towards  $\mathbf{a}(\theta_s, \phi_s)$ , the beam pattern becomes

$$\mathbf{b}_{MF}(\theta_s, \phi_s, \theta, \phi) = \mathbf{a}(\theta_s, \phi_s)^H \mathbf{a}(\theta, \phi) \quad (5)$$

Matched-filter beamforming passes a desired signal arriving from an a priori known spatial angle but rejects interferences or noise from all other possible angles. A narrow mainlobe and low sidelobes are desirable in the beamformer output.

### Construction #1:

$$\mathbf{b}_{MF} = \mathbf{b}_{MF}^{(3+1)} = (1 + \cos \psi)/2$$

which is independent of  $(\theta_s, \phi_s)$  and  $(\theta_i, \phi_i)$  per se, except through  $\psi$ . This constitutes a rotational invariance in the spherical coordinate; only the angular separation between the desired source's and the interference's arrival angles, not their absolute values, affects  $\mathbf{b}_{MF}^{(3+1)}$ .

### Construction #2:

$$\mathbf{b}_{MF} = \mathbf{b}_{MF}^{(3+0)}(\psi) = \cos \psi$$

which is also independent of  $(\theta_s, \phi_s)$  and  $(\theta_i, \phi_i)$  per se, except through  $\psi$ .  $\mathbf{b}_{MF}^{(3+1)}$  and  $\mathbf{b}_{MF}^{(3+0)}$  are plotted in Figure 1. The former has a single peak at  $\psi = 0$ , as expected; however, the latter suffers a  $\pi$ -ambiguity because interferences may come through at a spurious spatial peak at  $\psi = \pi$ .

### Construction #3:

$$\begin{aligned} \mathbf{b}_{MF} &= \mathbf{b}_{MF}^{(2+1)} = \frac{1}{2} + \frac{\cos \psi - \cos \theta_s \cos \theta}{2} \\ &= \frac{1}{2} + \frac{\sin \theta_s \sin \theta \cos(\phi_s - \phi)}{2} \end{aligned}$$

Unlike  $\mathbf{b}_{MF}^{(3+1)}$  and  $\mathbf{b}_{MF}^{(3+0)}$ ,  $\mathbf{b}_{MF}^{(2+1)}$  (plotted in Figure 2) is a function of  $\theta_s$  and  $\theta$ , in addition to  $\psi$ . There exists a rotational invariance only with respect to z-axis (i.e., the absolute values of the desired source's and interference's azimuth angles do not matter, only their difference). As  $\theta_s, \theta \in [0, \pi]$ , all trigonometric terms above are positive; thus,  $\mathbf{b}_{MF}^{(2+1)}$  has a minimum spatial response equal to 0.5. This means that interference may pass through from all angles regardless of the nominal arrival direction towards which  $\mathbf{b}_{MF}^{(2+1)}$  is pointed. Moreover,  $\cos(\phi_s - \phi)$  implies the existence of a spurious peak and thus a  $\pi$ -ambiguity in the azimuth angle. These properties render  $\mathbf{b}_{MF}^{(2+1)}$  a most unattractive matched-filter beamformer.

### Construction #4:

$$\mathbf{b}_{MF} = \mathbf{b}_{MF}^{(2+0)} = \cos \psi - \cos \theta_s \cos \theta = \sin \theta_s \sin \theta \cos(\phi_s - \phi_i)$$

Again,  $\mathbf{b}_{MF}^{(2+0)}$  (plotted in Figure 3) depends on  $\theta_s$  and  $\theta$  in addition to  $\psi$ ; and  $\mathbf{b}_{MF}^{(2+0)}$  is always positive. An incident signal with  $\psi = 0$  may be rejected, when  $\theta_s$  or  $\theta$  approaches 0 or  $\pi$ . Like  $\mathbf{b}_{MF}^{(2+1)}$ ,  $\mathbf{b}_{MF}^{(2+0)}$  offers little elevation maneuverability.  $\mathbf{b}_{MF}^{(2+0)}$  may be a useful in selective interference rejection only if all sources are known to impinge from  $\theta \approx 0$ . A  $\pi$ -ambiguity also exists; it arises from the  $\cos(\phi_s - \phi)$  in  $\mathbf{b}_{MF}^{(2+0)}$ .

**Conclusion:** The z-axis velocity hydrophone offers beamforming maneuverability in elevation.  $\mathbf{b}^{(2+1)}$  is useless as a matched-filter beamformer. Among the four vector-hydrophone constructions above, only the four-component vector hydrophone suffers no spurious peaks. The mainlobe in  $\mathbf{b}_{MF}^{(3+1)}(\psi)$  may be sharpened by deploying multiple four-component vector hydrophones in a spatially displaced array. The overall vector hydrophone array's spatial angular response equals the product between (1) the individual four-component vector hydrophone's spatial angular response, and (2) the spatial angular response of an array of omnidirectional sensors spaced in such a geometry [22].

## 2.2. Maximum-SINR Statistically Optimal Beamforming

The maximum-SINR beamformer aims to maximize the ratio of the desired signal's power over the combined power of all interference and noise. A high SINR is desirable over the widest range of azimuth and elevation angles. If (1) the additive noise is uncorrelated temporally and across hydrophones, and if (2) all signals and interferers and noise have zero cross-correlation, the data autocorrelation matrix  $\mathbf{R}_z$  of the input to the beamformer may be modeled as:

$$\mathbf{R}_z = \mathcal{P}_s \mathbf{a}_s \mathbf{a}_s^H + \mathcal{P}_n \mathbf{I} + \sum_{k=1}^K \mathcal{P}_k \mathbf{a}_k \mathbf{a}_k^H$$

where  $\mathcal{P}_s$  denotes the desired signal's power,  $\mathcal{P}_n$  symbolizes the noise power,  $\mathcal{P}_k$  refers to the  $k$ th interferer's power,  $\mathbf{a}_s$  represents the desired signal's steering vector, and  $\mathbf{a}_k$  symbolizes the  $k$ th interferer's steering vector, and  $K$  denotes the total number of interferers. The SINR equals:

$$\text{SINR} = \frac{\mathbf{w}^H (\mathcal{P}_s \mathbf{a}_s \mathbf{a}_s^H) \mathbf{w}}{\mathbf{w}^H (\mathbf{R}_z - \mathcal{P}_s \mathbf{a}_s \mathbf{a}_s^H) \mathbf{w}}$$

where  $\mathbf{w}$  refers to the beamforming weight vector. The  $\mathbf{w}$  that maximizes the SINR equals:

$$\mathbf{w}_{\text{SINR}}^{\circ} = \arg \mathbf{w}^{\max} \text{SINR} = \frac{\mathbf{R}_z^{-1} \mathbf{a}_s}{\mathbf{a}_s^H \mathbf{R}_z^{-1} \mathbf{a}_s}$$

Unlike  $\mathbf{w}_{\text{MF}}$ ,  $\mathbf{w}_{\text{SINR}}^{\circ}$  is a function of the collected data through  $\mathbf{R}_z$ . With  $K = 1$ , define  $\mathcal{P}_1 = \mathcal{P}_i$  and  $\mathbf{a}_k = \mathbf{a}_i$ . Using the relation

$$(\mathbf{A} + \mathbf{B}\mathbf{C}^H)^{-1} = \mathbf{A}^{-1} - \mathbf{A}^{-1}\mathbf{B}(\mathbf{I} + \mathbf{C}^H\mathbf{A}^{-1}\mathbf{B})^{-1}\mathbf{C}^H\mathbf{A}^{-1}$$

and setting  $\mathbf{A} = \mathcal{P}_n \mathbf{I}$  and  $\mathbf{B} = \mathbf{C} = \mathcal{P}_i \mathbf{a}_i$ , the maximum SINR ( $\text{SINR}^{\circ}$ ) becomes

$$\begin{aligned} \text{SINR}^{\circ} &= \frac{\mathcal{P}_s}{\mathcal{P}_n} \mathbf{a}_s^H \left( \mathbf{I} - \frac{\mathcal{P}_i \mathbf{a}_i \mathbf{a}_i^H}{1 + \frac{\mathcal{P}_i}{\mathcal{P}_n} \mathbf{a}_i^H \mathbf{a}_i} \right) \mathbf{a}_s \\ &= \frac{\mathcal{P}_s}{\mathcal{P}_n} \left( \|\mathbf{a}_s\|^2 - \frac{\mathcal{P}_i \|\mathbf{a}_i \mathbf{a}_s^H\|^2}{1 + \frac{\mathcal{P}_i}{\mathcal{P}_n} \|\mathbf{a}_i\|^2} \right) \end{aligned} \quad (6)$$

**Construction #1:**  $\|\mathbf{a}_s^{(3+1)}\|^2 = \|\mathbf{a}_i^{(3+1)}\|^2 = 2$ , and  $\|(\mathbf{a}_s^{(3+1)})^H \mathbf{a}_i^{(3+1)}\| = 1 + \cos \psi$ . Hence,

$$\text{SINR}^{\circ(3+1)} = \frac{\mathcal{P}_s}{\mathcal{P}_n} \left[ 2 - \frac{(1 + \cos \psi)^2}{2 + \left(\frac{\mathcal{P}_i}{\mathcal{P}_n}\right)^{-1}} \right]$$

which is independent of  $(\theta_s, \phi_s)$  and  $(\theta_i, \phi_i)$  per se, except through  $\psi$ . This constitutes a rotational invariance in the spherical coordinate. Only the angular separation between the desired source's and the interference's arrival angles, not their absolute values, affects  $\text{SINR}^{\circ(3+1)}$ .  $\frac{\text{SINR}^{\circ(3+1)}}{\mathcal{P}_s/\mathcal{P}_n}$

is plotted in Figure 4. Note that  $\text{SINR}^{\circ(3+1)}$  has a unique minimum at  $\psi = 0$  (when the desired source and the interference impinge from the same direction of arrival). The amplitude response's flat plateau peaks at  $\psi = \pi$ , when the desired source and the interference impinge from diametrically opposite directions of arrival.

**Construction #2:**  $\|\mathbf{a}_s^{(3+0)}\|^2 = \|\mathbf{a}_i^{(3+0)}\|^2 = 1$ , and  $\|(\mathbf{a}_s^{(3+0)})^H \mathbf{a}_i^{(3+0)}\| = \cos \psi$ . Hence,

$$\text{SINR}^{\circ(3+0)} = \frac{\mathcal{P}_s}{\mathcal{P}_n} \left[ 1 - \frac{\cos^2 \psi}{1 + \left(\frac{\mathcal{P}_i}{\mathcal{P}_n}\right)^{-1}} \right]$$

which is also independent of  $(\theta_s, \phi_s)$  and  $(\theta_i, \phi_i)$  per se, except through  $\psi$ .  $\frac{\text{SINR}^{\circ(3+0)}}{\mathcal{P}_s/\mathcal{P}_n}$  is plotted in Figure 5.

$\text{SINR}^{\circ(3+0)}$  has two minima at  $\psi = 0$  and  $\psi = \pi$ ; that is, when the desired source and the interference impinge from the same or from diametrically opposite directions of arrival.  $\text{SINR}^{\circ(3+0)}$  is maximum at  $\psi = \pi/2$  and  $\psi = 3\pi/2$ , when the desired source and the interference impinge from perpendicular directions of arrival. The double maxima and double minima mean that this vector-hydrophone construction, without a pressure hydrophone, suffers a  $180^\circ$  hemispherical ambiguity in  $\psi$ . This is because the absence of the pressure hydrophone means that acoustical dilation cannot be distinguished from acoustical compression.

**Construction #3:**  $\|\mathbf{a}_s^{(2+1)}\|^2 = 1 + \sin^2 \theta_s$ ,  $\|\mathbf{a}_i^{(2+1)}\|^2 = 1 + \sin^2 \theta_i$  and  $\|(\mathbf{a}_s^{(2+1)})^H \mathbf{a}_i^{(2+1)}\| = 1 + \sin \theta_s \sin \theta_i \cos(\phi_s - \phi_i) = 1 + \cos \psi - \cos \theta_s \cos \theta_i$ . Hence,

$$\begin{aligned} \text{SINR}^{\circ(2+1)} &= \frac{\mathcal{P}_s}{\mathcal{P}_n} \left[ (1 + \sin^2 \theta_s) - \frac{(1 + \cos \psi - \cos \theta_s \cos \theta_i)^2}{(1 + \sin^2 \theta_i) + \left(\frac{\mathcal{P}_i}{\mathcal{P}_n}\right)^{-1}} \right] \\ &= \frac{\mathcal{P}_s}{\mathcal{P}_n} \left[ (1 + \sin^2 \theta_s) - \frac{(1 + \sin \theta_s \sin \theta_i \cos(\phi_s - \phi_i))^2}{(1 + \sin^2 \theta_i) + \left(\frac{\mathcal{P}_i}{\mathcal{P}_n}\right)^{-1}} \right] \end{aligned}$$

Unlike  $\text{SINR}^{\circ(3+1)}$  and  $\text{SINR}^{\circ(3+0)}$ ,  $\text{SINR}^{\circ(2+1)}$  depends on  $\theta_s$  and  $\theta_i$ , in addition to  $\psi$ . There exists a rotational invariance only with respect to z-axis (i.e., the absolute values of the desired source's and interference's azimuth angles do not matter, only their difference).  $\frac{\text{SINR}^{\circ(2+1)}}{\mathcal{P}_s/\mathcal{P}_n}$  has the following properties:

(1) Nulls exist at  $\theta_s \approx 0$  and  $\theta_s \approx \pi$ , for all  $\theta_i$ ,  $\phi_s - \phi_i$  and  $\frac{\mathcal{P}_i}{\mathcal{P}_n}$ . Hence,  $\text{SINR}^{\circ(2+1)}$  is useful only if the interference is a priori known to impinge from near-horizontal arrival angles.

(2) When  $\frac{\mathcal{P}_i}{\mathcal{P}_n}$  is less than or roughly 0.1,  $\frac{\text{SINR}^{\circ(2+1)}}{\mathcal{P}_s/\mathcal{P}_n}$  depends only on  $\sin^2 \theta_s$  and roughly equals  $1 + \sin^2 \theta_s$ , which always exceeds or equal unity. Figure 6 shows that for  $\frac{\mathcal{P}_i}{\mathcal{P}_n} < 0.1$ , the SINR pattern is largely invariant with respect to  $\frac{\mathcal{P}_i}{\mathcal{P}_n}$  and  $|\phi_s - \phi_i|$ .

(3) For  $\frac{\mathcal{P}_i}{\mathcal{P}_n}$  equals to or exceeds unity and small  $|\phi_s - \phi_i|$ , two additional nulls appear at  $\theta_s \approx \theta_i$  and  $\theta_s \approx \pi - \theta_i$ . These two nulls, unlike those in (1), are no longer  $\pi$  radians apart. Hence, even if all sources are a priori known to impinge from one hemispherical side of the vector hydrophone, interference may still pass through unhindered when  $\frac{\mathcal{P}_i}{\mathcal{P}_n}$  exceeds about 0.2.

These above properties imply that  $\text{SINR}^{\circ(2+1)}$  works well only if the additive noise dominates the interference and only if the desired signal is known to impinge near-horizontally from one particular hemispherical side of the vector hydrophone. Note that when  $\theta_s = \theta_i = \pi/2$ ,  $\text{SINR}^{\circ(2+1)}$  is equivalent to  $\text{SINR}^{\circ(3+1)}$ . That is,  $\text{SINR}^{\circ(2+1)}$  and  $\text{SINR}^{\circ(3+1)}$  have the same response on the x-y plane.

**Construction #4:**  $\|\mathbf{a}_s^{(2+0)}\|^2 = \sin^2 \theta_s$ ,  $\|\mathbf{a}_i^{(2+0)}\|^2 = \sin^2 \theta_i$  and  $\|(\mathbf{a}_s^{(2+0)})^H \mathbf{a}_i^{(2+0)}\| = \sin \theta_s \sin \theta_i \cos(\phi_s - \phi_i) = \cos \psi - \cos \theta_s \cos \theta_i$ . Hence,

$$\begin{aligned} \text{SINR}^{o(2+0)} &= \frac{P_s}{P_n} \left[ \sin^2 \theta_s - \frac{\frac{P_i}{P_n} (\cos \psi - \cos \theta_s \cos \theta_i)^2}{1 + \left(\frac{P_i}{P_n} \sin^2 \theta_i\right)} \right] \\ &= \frac{P_s}{P_n} \sin^2 \theta_s \left[ 1 - \frac{\left(\frac{P_i}{P_n} \sin^2 \theta_i\right) \cos^2(\phi_s - \phi_i)}{1 + \left(\frac{P_i}{P_n} \sin^2 \theta_i\right)} \right] \end{aligned}$$

$\text{SINR}^{o(2+0)}$  relates to  $\theta_s$  as a linear function of  $\sin^2 \theta_s$ , but  $\text{SINR}^{o(2+0)}$  depends non-linearly on  $\frac{P_i}{P_n} \sin^2 \theta$  and  $\cos^2(\phi_s - \phi_i)$ . The non-linear dependencies are plotted in Figure 6, where the z-axis gives  $\frac{\text{SINR}^{o(2+0)}}{\frac{P_s}{P_n} \sin^2 \theta_s}$ .  $\text{SINR}^{o(2+0)}$  is still characterized by an azimuth rotational invariance. The nulls of  $\text{SINR}^{o(2+0)}$  lie at  $\phi_s = \phi_i$  and  $\phi_s = \phi_i + \pi$ , when  $\frac{P_i}{P_n} \sin^2 \theta_i \gg 0$  (i.e., when the x-y plane component of the interferer's power greatly exceeds noise power.) The extra null region at  $\phi_s = \phi_i + \pi$  arises from the absence of the pressure hydrophone like the case with  $\text{SINR}^{o(3+0)}$ , allowing interference to pass through unhindered. When  $\theta_s = \theta_i = \pi/2$ ,  $\text{SINR}^{o(2+0)}$  is equivalent to  $\text{SINR}^{o(3+0)}$ . That is,  $\text{SINR}^{o(2+0)}$  and  $\text{SINR}^{o(3+0)}$  have the same response on the x-y plane.

**Conclusion:** Only  $\text{SINR}^{o(3+1)}$  suffers no spurious null. If interference is a priori known to impinge from one particular hemispherical side of the vector hydrophone,  $\text{SINR}^{o(3+0)}$  will be equally usable as  $\text{SINR}^{o(3+1)}$ . Without the z-axis velocity hydrophone,  $\text{SINR}^{o(2+1)}$  and  $\text{SINR}^{o(2+0)}$  cannot reject near-vertical interference. The rather irregular beam pattern of  $\text{SINR}^{o(2+1)}$  renders it relatively useless.

### 3. OVERALL CONCLUSIONS

The four-component vector hydrophone offers a unimodal mainlobe in matched-filter beamforming and the broadest peak in maximum-SINR beamforming, with full maneuverability in elevation in addition to azimuth. However, if the incident sources are known to impinge from only one particular hemispherical side of the vector hydrophone,  $\text{SINR}^{o(2+0)}$  and  $\text{SINR}^{o(3+0)}$  also offer a unimodal mainlobe in matched-filter beamforming and the a broad peak in maximum-SINR beamforming.  $\text{SINR}^{o(2+0)}$  is especially useful for the case where the vertical oceanic acoustics need to be overlooked in order for the rectilinear model to be valid. In contrast,  $\text{SINR}^{o(2+1)}$  produces an essentially useless matched-filter beamforming pattern and an unreliable maximum-SINR beamforming pattern.

### 4. BIBLIOGRAPHY

- [1] C. B. Leslie, J. M. Kendall & J. L. Jones, "Hydrophone for Measuring Particle Velocity", *J. Acoustical Soc. of America*, vol. 28, no. 4, pp. 711-715, July 1956.
- [4] E. L. Alder, "Acoustic Beamforming with Vector Sensors," *IEEE Intl. Symp. Spread Spectrum Applications*, 1987.
- [6] B. D. Van Veen & K. M. Buckley, "Beamforming: A versatile Approach to Spatial Filtering," *IEEE Acoustics, Speech & Signal Processing Magazine*, pp. 4-24, Apr. 1988.
- [7] G. L. D'Spain, W. S. Hodgkiss & G. L. Edmonds, "The Simultaneous Measurement of Infrasonic Acoustic Particle Velocity and Acoustic Pressure in the Ocean by Freely Drifting Swallow Floats," *IEEE J. Oceanic Engineering*, vol. 16, pp. 195-207, 1991.
- [8] G. L. D'Spain, W. S. Hodgkiss & G. L. Edmonds, "Energetics of the Deep Ocean's Infrasonic Sound Field," *J. Acoustical Soc. America*, pp. 1134-1158, March 1991.
- [9] J. C. Nickles, G. Edmonds, R. Harriss & etc., "A Vertical Array of Directional Acoustic Sensors," *IEEE Oceans Conf.*, pp. 340-345, 1992.
- [10] G. L. D'Spain, W. S. Hodgkiss, G. L. Edmonds, J. C. Nickles, F. H. Fisher, R. A. Harris & etc., "Initial Analysis of the Data from the Vertical DIFAR Array," *IEEE Oceans Conf.*, pp. 346-351, 1992.
- [11] H. Cox & R. M. Zeskind, "Adaptive Cardioid Processing," *26th Asilomar Conference* pp. 1058-1061, 1992.
- [12] V. A. Shchurov, V. I. Ilyichev & V. P. Kuleshov, "Ambient Noise Energy Motion in the Near Surface Layer in Ocean Wave-Guide," *Journal de Physique*, vol. 4, no. 5, part 2, pp. 1273-1276, May 1994.
- [13] A. Nehorai & E. Paldi, "Acoustic Vector Sensor Array Processing," *IEEE Trans. Signal Processing*, vol. 42, no. 9, pp. 2481-2491, Sep. 1994.
- [14] M. Hawkes & A. Nehorai, "Bearing Estimation with Acoustic Vector-Sensor Arrays," *Acoustic Velocity Focused Workshop*, pp. 345-348, 1995.
- [15] M. Hawkes & A. Nehorai, "Hull-Mounted Acoustic Vector-Sensor Arrays," *Asilomar Conf.*, pp. 1046-1050, 1995.
- [16] M. Hawkes & A. Nehorai, "Surface Mounted Acoustic Vector-Sensor Array Processing," *IEEE Intl. Conf. Acoustic, Speech & Signal Processing*, pp. 3170-3173, 1996.
- [17] B. Hochwald & A. Nehorai, "Identifiability in Array Processing Models with Vector-Sensor Applications," *IEEE Trans. Signal Processing*, pp. 83-95, Jan. 1996.
- [18] M. J. Berliner & J. F. Lindberg, *Acoustic Particle Velocity Sensors: Design, Performance & Applications*, Woodbury, N.Y.:AIP Press, 1996.
- [19] K. T. Wong & M. D. Zoltowski, "Orthogonal Velocity-Hydrophone ESPRIT for Sonar Source Localization," *MTS/IEEE Oceans Conference*, vol. 3, pp. 1307-1312, 1996.
- [20] K. T. Wong & M. D. Zoltowski, "Closed-form Underwater Acoustic Direction-Finding with Arbitrarily Spaced Vector-Hydrophones at Unknown Locations," *IEEE J. of Oceanic Engineering*, pp. 566-575, July 1997.
- [21] K. T. Wong & M. D. Zoltowski, "Extended-Aperture Underwater Acoustic Multisource Azimuth/Elevation Direction-Finding Using Uniformly But Sparsely Spaced Vector Hydrophones," *IEEE J. Oceanic Engineering*, pp. 659-672, October 1997.
- [22] R. C. Hansen, *Phased Array Antennas*, Wiley, 1998.
- [23] M. Hawkes & A. Nehorai, "Acoustic Vector-Sensor Beamforming and Capon Direction Estimation," *IEEE Trans. Signal Processing* pp. 2291-2304, Sept. 1998.
- [24] M. Hawkes & A. Nehorai, "Effects of Sensor Placement on Acoustic Vector-Sensor Array Performance," *IEEE J. Oceanic Engineering*, vol. 24, no. 1, pp. 33-40, Jan. 1999.
- [25] K. T. Wong, "Adaptive Source Localization & Blind Beamforming for Underwater Acoustic Wideband Fast Frequency-Hop Signals of Unknown Hop Sequences & Unknown Arrival Angles Using a Vector-Hydrophone," *IEEE Wireless Communications & Networking Conference* 1999.
- [26] K. T. Wong & M. D. Zoltowski, "Root-MUSIC-Based Azimuth-Elevation Angle-of-Arrival Estimation with Uniformly Spaced but Arbitrarily Oriented Velocity Hydrophones," *IEEE Trans. Signal Processing*, pp. 3250-3260, Dec. 1999.
- [27] K. T. Wong & M. D. Zoltowski, "Self-Initiating Velocity-Field Beam-space MUSIC for Underwater Acoustic Direction-Finding Using Irregularly Spaced Vector-Hydrophones," *J. Oceanic Engineering*, April 2000.

Figure 1: Matched-Filter Beam Pattern for Vector-Hydrophone Constructions #1 & #2:

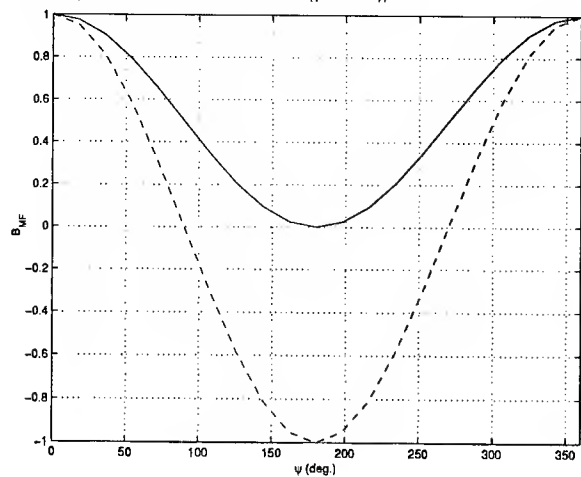


Figure 4: Maximum-SINR Beam Pattern for Vector-Hydrophone Construction #1:

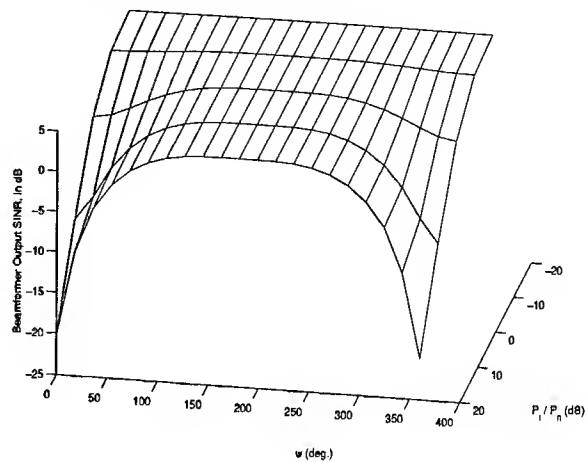


Figure 2: Matched-Filter Beam Pattern for Vector-Hydrophone Construction #3:

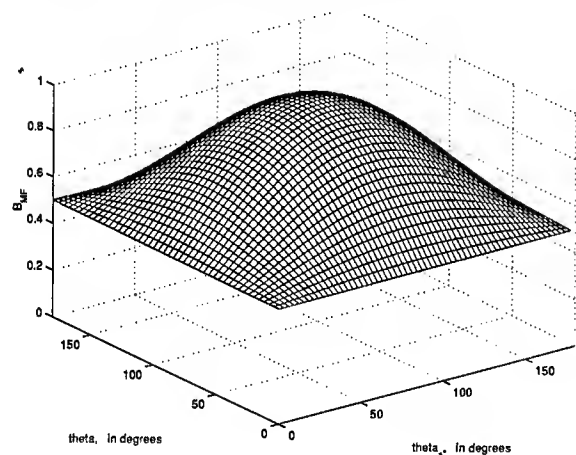


Figure 5: Maximum-SINR Beam Pattern for Vector-Hydrophone Construction #2:

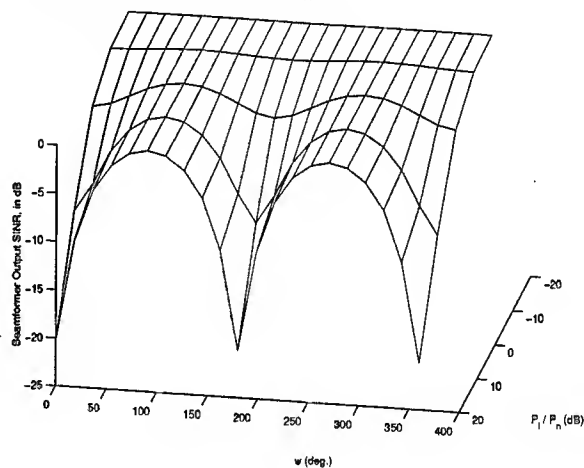


Figure 3: Matched-Filter Beam Pattern for Vector-Hydrophone Construction #4:

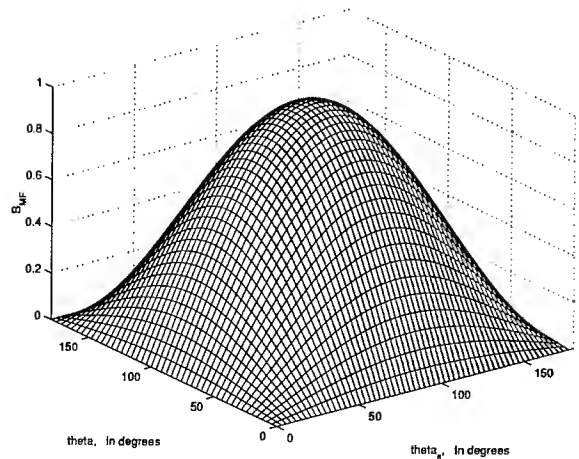
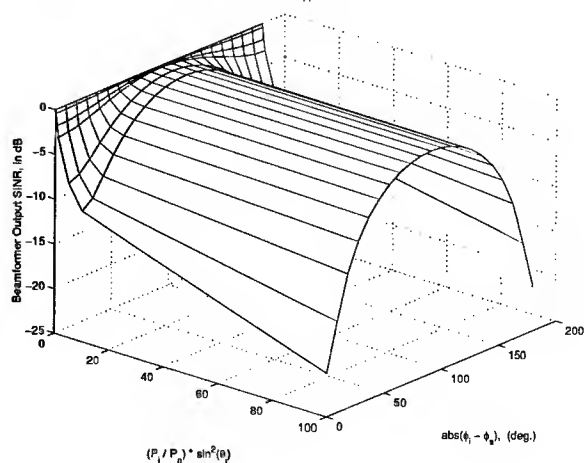


Figure 6: Maximum-SINR Beam Pattern for Vector-Hydrophone Construction #4:



# Author Index

## A

Abdi, A. ....58  
 Abdul-Jauwad, S. H. ....544  
 Abed-Meraim, K. ....90, 334  
 Abramovich, Y. I. ....99  
 Adams, V. ....495  
 Adnet, C. ....315  
 Akmouche, W. ....48  
 Alengrin, G. ....276  
 Amin, M. G. ....23, 467, 529, 682  
 Andrieu, C. ....6, 131, 405  
 Aouada, S. ....43  
 Attallah, S. ....90  
 Au, K. K. ....166  
 Aviyente, S. ....569

## B

Badidi, L. ....434  
 Barkat, B. ....262, 594  
 Bar-Ness, Y. ....15  
 Barrère, J. ....319  
 Barroso, V. ....243, 329, 607, 727  
 Bastani, M. H. ....444  
 Batalama, S. N. ....524, 677  
 Bech, M. M. ....373  
 Belouchrani, A. ....43, 306, 621  
 Bennink, D. ....598  
 Besson, O. ....424  
 Bi, Z. ....229  
 Biguesh, M. ....444  
 Biracree, S. ....645  
 Birkelund, Y. ....640  
 Blum, R. S. ....1  
 Boashash, B. ....559, 564, 584  
 Böhme, J. F. ....68  
 Borloz, B. ....349  
 Bouillaut, L. ....702  
 Bousbia-Salah, H. ....306  
 Braunreiter, D. C. ....472  
 Brcich, R. ....94, 448  
 Brown, C. L. ....594  
 Bugallo, M. F. ....10

## C

Cardoso, J. ....359  
 Carson, H. ....564  
 Casas, R. A. ....645  
 Cassabaum, M. L. ....472  
 Castanie, F. ....176  
 Castedo, L. ....10  
 Cavassilas, J-F ....349  
 Chabriel, G. ....319  
 Champagne, B. ....444

Chang, E. ....717  
 Chaparro, L. F. ....519  
 Chapman, N. R. ....635  
 Chen, C. ....104  
 Chen, C. ....216  
 Chen, C. ....354  
 Chen, H.-W. ....472  
 Chen, R. ....239  
 Chen, V. C. ....463  
 Cheng, H. ....490, 549  
 Chi, C-Y ....216, 354  
 Chi, H. ....732  
 Chkeif, A. ....90  
 Chung, P. J. ....68  
 Ciblat, P. ....171  
 Cichocki, A. ....626, 631  
 Cohen, G. ....396  
 Cohen, L. ....410, 485, 589  
 Colas, M. ....655  
 Collings, I. B. ....151  
 Comon, P. ....181, 206  
 Cowper, M. R. ....267, 296

## D

Dasgupta, S. ....211  
 Davidson, K. L. ....574  
 Davis, L. M. ....151  
 Delaunay, G. ....655  
 Derras, B. ....621  
 Dietrich, F. ....311  
 Ding, Z. ....616  
 Dizaji, R. M. ....378, 635  
 Djurovic, Z. ....243  
 Doherty, J. F. ....191  
 Doroslovački, M. ....534  
 Doucet, A. ....6, 131  
 Durrani, T. ....368  
 Duverdier, A. ....196

## E

El-Jaroudi, A. ....603  
 Endres, T. J. ....645  
 Enescu, M. ....301  
 Escudero, C. J. ....10  
 Evans, R. J. ....151

## F

Fernández-Rubio, J. A. ....668  
 Ferrari, A. ....276  
 Fiedler, R. ....480  
 Filipowicz, S. F. ....631

Fishler, E. ....86, 225  
 Francos, J. M. ....118, 396  
 Friedmann, J. ....225  
 Frikel, M. ....141

## G

Galleani, L. ....410, 589  
 Galy, J. ....315, 655  
 Gelle, G. ....655  
 Gershman, A. B. ....68, 78, 424, 467  
 Gharieb, R. R. ....626, 631  
 Ghogho, M. ....368  
 Giannakis, G. B. ....171  
 Gillespie, W. ....717  
 Goldstein, J. S. ....514  
 Golikov, V. S. ....83  
 Gomes, J. ....727  
 Gray, D. ....712  
 Green, R. A. ....664  
 Grellier, O. ....206  
 Groutage, D. ....598  
 Gu, Q. ....510

## H

Han, D-S ....439  
 Hanssen, A. ....391, 539, 640, 650  
 Hasan, A. A. ....127  
 Hasan, M. A. ....127  
 Hatzinakos, D. ....166  
 Hedges, R. A. ....252  
 Hillery, W. J. ....18  
 Hinich, M. J. ....281  
 Hlawatsch, F. ....554  
 Hong, Z. ....221  
 Horita, Y. ....626  
 Hua, Y. ....90  
 Huang, W. ....191  
 Hulyalkar, S. ....645  
 Hussain, Z. M. ....559

## I

Iglesia, D. I. ....10  
 Igual, J. ....364  
 Ilow, J. ....505

## J

Jansen, R. ....480  
 Jones, D. L. ....53, 429

## K

Kadtke, J. B. ....722  
 Kamran, Z. M. ....334

# Author Index

Karan, M. ....712  
 Kassam, S. A. ....161, 324  
 Kaveh, M. ....58  
 Kerherve, E. ....48  
 Khan, I. ....534  
 Khène, M. F. ....544  
 Kirilin, R. L. ....373, 378, 635  
 Kirsteins, I. P. ....286  
 Kliger, M. A. ....118  
 Knight, A. J. ....707  
 Koivunen, V. ....301  
 Kosanović, B. ....534  
 Kovacevic, B. ....243  
 Kozick, R. J. ....73, 419  
 Krauss, T. P. ....18  
 Kraut, S. ....113  
 Krolik, J. ....113  
 Krongold, B. S. ....53  
 Kuzminskiy, A. M. ....156

## L

Lacaze, B. ....196  
 Lagunas, M. A. ....248, 383  
 Larsen, Y. ....539  
 Leduc, J. ....136  
 Lee, J. ....660  
 Lefkadtis, V. ....387  
 Lennartsson, R. K. ....281, 722  
 Leyman, A. R. ....334  
 Li, B-W ....216  
 Li, H. ....229  
 Li, J. ....229  
 Li, J. ....476  
 Liang, J. ....616  
 Lindsey, A. R. ....529  
 Ling, H. ....476  
 Liu, D. ....229  
 Liu, J. S. ....239  
 Liu, P. ....33, 692  
 López-Valcarce, R. ....211  
 Loubaton, P. ....171  
 Loughlin, P. J. ....574  
 Luschi, C. ....156, 201

## M

Mancuso, V. ....682  
 Manikas, A. ....387  
 Mansour, A. ....63  
 Martin, C. ....186  
 Marx, D. ....717  
 Matz, G. ....554  
 McLaughlin, S. ....281, 296  
 Meddeb, S. ....176

Medley, M. J. ....524  
 Mesbah, M. ....564  
 Messer, H. ....86, 225, 453  
 Mestre, X. ....248  
 Mizuguchi, Y. ....28  
 Moon, S. ....439  
 Moreau, E. ....339, 344  
 Morelande, M. R. ....262  
 Mulgrew, B. ....201, 267, 296  
 Murai, T. ....626  
 Murphy, C. D. ....161  
 Murphy, C. D. ....697  
 Murrow, D. J. ....272  
 Myrick, W. L. ....514

## N

Nelson, D. J. ....400  
 Nelson, M. ....717  
 Nickel, R. M. ....612  
 Nossek, J. ....141

## O

Ohnishi, N. ....63  
 Oliveira, P. M. ....607  
 Ottersten, B. ....186  
 Oxley, M. E. ....257

## P

Pagès-Zamora, A. ....248  
 Papandreou-Suppappola, A. ....579  
 Pareja, F. C. ....83  
 Peake, M. ....712  
 Pelin, P. ....94, 448  
 Péntek, Á. ....722  
 Pérez, J.-M. ....405  
 Perez-Neira, A. I. ....383  
 Persson, L. ....281  
 Pesavento, M. ....78, 467  
 Petrochilos, N. ....181  
 Petropulu, A. P. ....495  
 Pitton, J. W. ....108  
 Psaromiligkos, I. N. ....677  
 Putney, A. ....717

## Q

Quinquis, A. ....48

## R

Rabideau, D. J. ....234  
 Radouane, L. ....434

Rangaswamy, M. ....286  
 Rao, A. M. ....429  
 Redfern, A. J. ....38  
 Reid, T. F. ....257  
 Rey, F. ....687  
 Riba, J. ....687  
 Rickard, S. ....311  
 Riddle, J. G. ....472  
 Robinson, J. W. C. ....281

## S

Sadler, B. M. ....73, 419  
 Salberg, A-B ....650  
 Samaras, K. ....156  
 Samuel, A. A. ....472  
 Schaffer, T. A. ....645  
 Schmidbauer, A. ....673  
 Schmitt, H. A. ....472  
 Scholl, J. F. ....472  
 Seco, G. ....668  
 Serpedin, E. ....171  
 Sethu, H. ....500  
 Shah, A. ....645  
 Sidahmed, M. ....702  
 Silverstein, S. D. ....291  
 Solomon, I. S. D. ....707  
 Spasojević, P. ....146  
 Spencer, N. K. ....99  
 Stoica, P. ....229, 424  
 Strauch, P. ....156  
 Strolle, C. H. ....645  
 Sucic, V. ....584  
 Suleesathira, R. ....519  
 Suppappola, S. B. ....579  
 Suter, B. W. ....252, 257  
 Swami, A. ....368  
 Swindlehurst, A. L. ....668

## T

Tabrikian, J. ....453  
 Thaiupathump, T. ....161  
 Thirion-Moreau, N. ....344  
 Thomson, D. J. ....414  
 Tourneret, J. Y. ....176, 196  
 Touzni, A. ....6  
 Trzynadlowski, A. M. ....373

## U

Unsworth, C. P. ....296  
 Utschick, W. ....141

## Author Index

### V

Valaee, S. ....444  
Vázquez, G. ....687  
Vergara, L. ....364

### W

Wang, J. ....378  
Wang, K. ....15  
Wang, X. ....146, 239  
Warman, K. ....717  
Wei, D. ....123, 490, 549  
Williams, W. J. ....569, 612  
Wohlberg, B. ....458  
Wong, K. T. ....732  
Wu, H-T .....104

### X

Xavier, J. ....329  
Xerri, B. ....349  
Xiang, L. ....324  
Xiong, P. ....524  
Xu, Z. ....33, 692

### Y

Yang, K. ....23, 28  
Yang, X. ....495  
Yu, K-B .....272  
Yun, D-H .....439

### Z

Zhang, Y. ....23, 28, 682  
Zhang, Y. ....324  
Zhang, Y. ....1  
Zhao, L. ....529  
Zhao, R. ....510  
Zheng, B. ....221  
Zhou, G. T. ....38  
Zhou, Y. ....500  
Zoltowski, M. D. ....18, 514  
Zoubir, A. M. ....94, 262, 448, 594  
Zweig, G. ....458